OM of: US-09-303-518D-127 to: A_Geneseq_032802:* out_format : pfs

Date: Jun 30, 2002 6:40 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09303518/runat_28062002_142712_4291/app_query.fasta_1.23501
-DB=A_Geneseq_032802 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09303518_@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-303-518D-127
Query length: 1344
Database: A_Geneseq_032802:*
Database sequences: 747574
Database length: 111073796
Search time (sec): 627.340000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation ..		
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38562	+	2208.00	4131.43	1.3e-222	44			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38561	+	2177.00	4073.01	2.3e-219	44			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38564	+	2148.00	4018.36	2.5e-216	44			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38563	+	1511.00	2821.82	1.6e-149	32			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38560	+	739.00	1375.49	1.2e-68	15			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34439	+	636.00	1168.77	1.3e-57	45			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34318	+	636.00	1168.69	1.3e-57	45			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75273	-	563.00	1047.11	3.0e-50	12			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75272	-	536.00	996.23	2.0e-47	12			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75271	-	487.50	904.93	2.5e-42	11			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY82082	+	447.50	813.11	8.3e-38	46			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY35375	+	447.50	813.06	8.3e-38	46			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY37553	+	432.00	784.06	3.5e-36	46			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34467	+	159.50	270.99	1.4e-07	44			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34343	+	159.50	270.77	1.4e-07	45			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59813	+	134.50	213.89	9.0e-05	10			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826	+	134.50	208.33	0.0001	16			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59817	+	132.00	209.39	0.0002	99			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04998	+	131.00	218.87	0.0001	38			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59827	+	128.00	196.25	0.0006	15			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB27242	-	122.50	198.21	0.0012	57			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB94220	-	122.00	195.91	0.0014	63			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03731	+	121.00	193.00	0.0019	69			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23389	-	113.50	172.38	0.0158	11			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU39645	-	112.00	184.72	0.0115	33			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY31745	+	110.00	178.06	0.0212	43			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU35592	+	109.50	176.39	0.0247	45			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG04814	-	108.50	168.89	0.0405	72			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB59201	-	107.00	165.79	0.0589	74			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB64198	-	107.00	151.71	0.1110	24			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23390	-	106.50	159.04	0.0863	12			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG90820	+	106.50	159.04	0.0863	12			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04955	+	105.50	166.13	0.0733	57			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW89449	+	105.00	170.38	0.0655	37			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW89448	+	105.00	168.50	0.0713	43			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW13502	+	104.50	151.57	0.1653	16			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:AAP93285	+	104.50	144.03	0.2321	30			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU34397	+	104.00	166.67	0.0905	43			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW18663	+	103.50	167.08	0.0961	38			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB15935	+	103.00	158.34	0.1541	74			
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG00972	-	101.50	134.44	0.5723				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG07375	-	101.50	134.13	0.5804				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG21944	-	101.50	134.10	0.5812				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR59843	-	101.00	164.99	0.1563				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG80035	+	101.00	159.12	0.2037				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG80043	+	101.00	159.12	0.2037				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG21954	-	100.50	149.11	0.3457				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR57141	-	100.50	141.47	0.4877				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB58723	-	99.50	140.77	0.5888				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG06375	-	99.00	135.84	0.7952				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM40851	+	98.50	143.00	0.6230				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG22380	+	98.50	135.49	0.8739				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY95559	-	98.50	133.57	0.9526				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY95556	-	98.50	132.35	1.01				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB27775	+	98.00	158.96	0.3283				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB32946	+	98.00	158.96	0.3283				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB18420	+	98.00	158.96	0.3283				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM53747	+	98.00	158.96	0.3283				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM66130	+	98.00	158.96	0.3283				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM14001	+	98.00	158.96	0.3283				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM26407	+	98.00	158.96	0.3283				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM01742	+	98.00	158.96	0.3283				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB27664	+	98.00	157.55	0.3499				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB28209	+	98.00	157.55	0.3499				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB32834	+	98.00	157.55	0.3499				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB33384	+	98.00	157.55	0.3499				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB18316	+	98.00	157.55	0.3499				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB18843	+	98.00	157.55	0.3499				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM53640	+	98.00	157.55	0.3499				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM54169	+	98.00	157.55	0.3499				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM66020	+	98.00	157.55	0.3499				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM66563	+	98.00	157.55	0.3499				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM13889	+	98.00	157.55	0.3499				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM14436	+	98.00	157.55	0.3499				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM26296	+	98.00	157.55	0.3499				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM26849	+	98.00	157.55	0.3499				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM01632	+	98.00	157.55	0.3499				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM02163	+	98.00	157.55	0.3499				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37266	+	98.00	155.31	0.3871				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU31533	-	98.00	154.90	0.3942				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG22379	+	98.00	139.95	0.7731				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG22375	+	98.00	139.20	0.7995				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM39065	+	98.00	138.29	0.8331				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG22378	+	98.00	137.78	0.8525				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB57783	-	97.50	149.41	0.5460				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW76719	+	97.50	139.59	0.8498				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW76720	+	97.50	139.59	0.8498				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:AAR92115	-	97.00	157.45	0.4110				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB65656	-	97.00	135.58	1.10				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG06375	+	97.00	132.07	1.29				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB86943	+	96.50	151.20	0.5893				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU33726	+	96.00	146.93	0.7723				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU36550	+	96.00	146.86	0.7749				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59824	+	96.00	135.85	1.27				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG08467	-	95.00	149.71	0.7973				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG82252	+	95.00	149.43	0.8072				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY00243	+	95.00	145.37	0.9692				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY28654	-	95.00	142.30	1.11				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB41835	-	95.00	141.84	1.14				
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59814	+	95.00	139.39	1.27				

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38562

seq_documentation_block:
ID AAY38562 standard; Protein; 447 AA.
XX
AC AAY38562;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria meningitidis strain A antigen encoded by ORF22.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

```
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12027.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 123; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   447 AA;


alignment_scores:
            Quality: 2208.00              Length:    447
              Ratio:  5.088                Gaps:      0
 Percent Similarity:  97.092   Percent Identity: 97.092

alignment_block:
 US-09-303-518D-127 x AAY38562   ..

 Align seg 1/1  to: AAY38562  from: 1  to: 447

       1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProGl 17

      51 GCAAGTCATTTATGACGGGCCCGTCATTACCGAAGTCGCGTTGCTTGGCG 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 uGlnValIleTyrAspGlyProValIleThrGluValAlaLeuLeuGlyG 34

     101 AAGAATATGCCGGTATGCGCCCCTNGATGAAAGTCAAGGAAGGCGATGCC 150
         ||||||||||||||||||||||   ||||||||||||||||||||||||
      34 luGluTyrAlaGlyMetArgPro***MetLysValLysGluGlyAspAla 50

     151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGNATCCGGGCGTGGT 200
         ||||||||||||||||||||||||||||||||||   ||||||||||||
      51 ValLysLysGlyGlnValLeuPheGluAspLysLys***ProGlyValVa 67

     201 GTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCATCCATCGCGGCGAAA 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 lPheThrAlaProValSerGlyLysIleAlaAlaIleHisArgGlyGluL 84

     251 AGCGCGTACTTCAGTCGGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 ysArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluIle 100

     301 GAGTTCGAACGCTACGCGCCCGAAGCGTTGGCAAACTTAAGCGGCGANGA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||  ||
     101 GluPheGluArgTyrAlaProGluAlaLeuAlaAsnLeuSerGly***Gl 117

     351 ANTNNGNNGCAATCTGATCCAATCCGGTTTGTGGACTGCGCTGCGTANCC 400
         |         ||||||||||||||||||||||||||||||||||   |
     117 u*********AsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArg***A 134

     401 GTCCGTTCAGCAAAATCCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTC 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150

     451 GTCAATGCGATGGACACCAATCCGCTNGCGGCAGACCCTGTGGTTGTGAT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProValValValIl 167

     501 CAAAGAAGCCGNCGANGATTTCAGACGANGTNTGCTGGTATTGAGCCGTT 550
         ||||||||||      ||||||||||||      ||||||||||||||||
     167 eLysGluAla******AspPheArgArg******LeuValLeuSerArgL 184

     551 TGACCGAGCGTAAAATCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCG 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 euThrGluArgLysIleHisValCysLysAlaAlaGlyAlaAspValPro 200

     601 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCC 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyGlyProHisPr 217

     651 GGCCGGTTTGAGTGGCACGCACATTCATTTCATTGAGCCGGTCGGTGCAA 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAlaA 234

     701 ACAAAACCGTTTGGACCATCAATTATCAAGATGTAATTGCCATCGGACGT 750
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 snLysThrValTrpThrIleAsnTyrGlnAspValIleAlaIleGlyArg 250

     751 TTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTGATTGCTTTGGGTGG 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 LeuPheAlaThrGlyArgLeuAsnThrGluArgValIleAlaLeuGlyGl 267

     801 TTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTTGGGTGCGAAAG 850
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 ySerGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284

     851 TATCGCAAATTACTGCGGGCGAATTGGTTGACGCAGACAACCGCGTGATT 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     284 alSerGlnIleThrAlaGlyGluLeuValAspAlaAspAsnArgValIle 300

     901 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTT 950
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 SerGlySerValLeuAsnGlyAlaIleThrGlnGlyAlaHisAspTyrLe 317

     951 GGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAG 1000
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     317 uGlyArgTyrHisAsnGlnIleSerValIleGluGluGlyArgSerLysG 334

    1001 AGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGT 1050
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     334 luLeuPheGlyTrpValAlaProGlnProAspLysTyrSerIleThrArg 350

    1051 ACGACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTTCACGACAGC 1100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 ThrThrLeuGlyHisPheLeuLysAsnLysLeuPheLysPheThrThrAl 367

    1101 CGTCAACGGTGGCGACCGCGCCATGGTGCCGATTGGTACTTACGAGCGCG 1150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTyrGluArgV 384

    1151 TAATGCCGCTAGACATCCTGCCTACCCTGCTTTTGCGCGATTTAATCGTC 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
     384 alMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIleVal 400

    1201 GGCGATACCGACAGCGCGCAAGCATTGGGTTGCTTGGAATTGGACGAAGA 1250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluGl 417

    1251 AGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATANGGCC 1300
         ||||||||||||||||||||||||||||||||||||||||||||   ||||
     417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGlu***GlyP 434

    1301 CGCTGTTGCGTAAGGTGCTGGAAACCNTTGAGAAGGAAGGC 1341
         |||||||||||||||||||||||||   ||||||||||||
     434 roLeuLeuArgLysValLeuGluThr***GluLysGluGly 447

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38561

seq_documentation_block:
ID   AAY38561 standard; Protein; 447 AA.
XX
AC   AAY38561;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis antigen encoded by ORF22.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12026.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 123; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   447 AA;


alignment_scores:
            Quality: 2177.00              Length:    447
              Ratio:  4.982                Gaps:      0
 Percent Similarity:  97.763   Percent Identity: 94.855
```

```
alignment_block:
 US-09-303-518D-127 x AAY38561   ..

 Align seg 1/1  to: AAY38561  from: 1  to: 447

       1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProGl 17

      51 GCAAGTCATTTATGACGGGCCCGTCATTACCGAAGTCGCGTTGCTTGGCG 100
         ||||::::::|||||||||||||:::|||||||||||||||||||||||||
      17 uGlnAlaValTyrAspGlyProAlaIleThrGluValAlaLeuLeuGlyG 34

     101 AAGAATATGCCGGTATGCGCCCCTNGATGAAAGTCAAGGAAGGCGATGCC 150
         |||||||||||||||||||||||:::|||||||||||||||||||||||||
      34 luGluTyrAlaGlyMetArgProSerMetLysValLysGluGlyAspAla 50

     151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGNATCCGGGCGTGGT 200
         |||||||||||||||||||||||||||||||||   ||||||||||||
      51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67

     201 GTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCATCCATCGCGGCGAAA 250
         ||||||||||||:::|||||||||||||||||||||||||||||||||||
      67 lPheThrAlaProAlaSerGlyLysIleAlaAlaIleHisArgGlyGluL 84

     251 AGCGCGTACTTCAGTCGGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 ysArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluIle 100

     301 GAGTTCGAACGCTACGCGCCCGAAGCGTTGGCAAACTTAAGCGGCGANGA 350
         ||||||||||||||||||||||||||||||||||||||||||||   ||
     101 GluPheGluArgTyrAlaProGluAlaLeuAlaAsnLeuSerGlyGluGl 117

     351 ANTNNGNNGCAATCTGATCCAATCCGGTTTGTGGACTGCGCTGCGTANCC 400
         |         ||||||||||||||||||||||||||||||||||||:::|
     117 uValArgArgAsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArgThrA 134

     401 GTCCGTTCAGCAAAATCCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTC 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150

     451 GTCAATGCGATGGACACCAATCCGCTNGCGGCAGACCCTGTGGTTGTGAT 500
         ||||||||||||||||||||||||||||||||||:::|||:::||
     151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProThrValIleIl 167

     501 CAAAGAAGCCGNCGANGATTTCAGACGANGTNTGCTGGTATTGAGCCGTT 550
         |||||||||:::   ||||||:::|||      ||||||||||||||||
     167 eLysGluAlaAlaGluAspPheLysArgGlyLeuLeuValLeuSerArgL 184

     551 TGACCGAGCGTAAAATCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCG 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 euThrGluArgLysIleHisValCysLysAlaAlaGlyAlaAspValPro 200

     601 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCC 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyGlyProHisPr 217

     651 GGCCGGTTTGAGTGGCACGCACATTCATTTCATTGAGCCGGTCGGTGCAA 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAlaA 234

     701 ACAAAACCGTTTGGACCATCAATTATCAAGATGTAATTGCCATCGGACGT 750
         ||||||||||||||||||||||||||||||||||:::|||||||||
     234 snLysThrValTrpThrIleAsnTyrGlnAspValIleThrIleGlyArg 250

     751 TTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTGATTGCTTTGGGTGG 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 LeuPheAlaThrGlyArgLeuAsnThrGluArgValIleAlaLeuGlyGl 267

     801 TTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTTGGGTGCGAAAG 850
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 ySerGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284
```

```
 851 TATCGCAAATTACTGCGGGCGAATTGGTTGACGCAGACAACCGCGTGATT 900
     ||||||||||||||||||||||||||||||:::|||||||||||||||||
 284 alSerGlnIleThrAlaGlyGluLeuValAspThrAspAsnArgValIle 300

 901 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTT 950
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 301 SerGlySerValLeuAsnGlyAlaIleThrGlnGlyAlaHisAspTyrLe 317

 951 GGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAG 1000
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 317 uGlyArgTyrHisAsnGlnIleSerValIleGluGluGlyArgSerLysG 334

1001 AGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGT 1050
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 334 luLeuPheGlyTrpValAlaProGlnProAspLysTyrSerIleThrArg 350

1051 ACGACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTTCACGACAGC 1100
     ||||||||||||||||||||||||||||||||||||||:::|||||||||
 351 ThrThrLeuGlyHisPheLeuLysAsnLysLeuPheLysPheAsnThrAl 367

1101 CGTCAACGGTGGCGACCGCGCCATGGTGCCGATTGGTACTTACGAGCGCG 1150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTyrGluArgV 384

1151 TAATGCCGCTAGACATCCTGCCTACCCTGCTTTTGCGCGATTTAATCGTC 1200
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 384 alMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIleVal 400

1201 GGCGATACCGACAGCGCGCAAGCATTGGGTTGCTTGGAATTGGACGAAGA 1250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluGl 417

1251 AGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATANGGCC 1300
     ||||||||||||||||||||||||||||||||||||||||||   ||||
 417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrGlyP 434

1301 CGCTGTTGCGTAAGGTGCTGGAAACCNTTGAGAAGGAAGGC 1341
     |||||||||||||||||||||||||   ||||||||||||
 434 roLeuLeuArgLysValLeuGluThrIleGluLysGluGly 447
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38564

seq_documentation_block:

```
ID   AAY38564 standard; Protein; 447 AA.
XX
AC   AAY38564;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria gonorrhoeae antigen encoded by ORF22.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12028.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 125; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   447 AA;
```

alignment_scores:

```
           Quality: 2148.00              Length:    447
             Ratio:  4.927                Gaps:      0
 Percent Similarity:  97.539   Percent Identity: 93.289
```

alignment_block:

US-09-303-518D-127 x AAY38564 ..

Align seg 1/1 to: AAY38564 from: 1 to: 447

```
   1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGA 50
     ||||||||||||||||||||||||||||||||||||||||||||||||||
   1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProGl 17

  51 GCAAGTCATTTATGACGGGCCCGTCATTACCGAAGTCGCGTTGCTTGGCG 100
     ||||||||||||||||||||||:::|||||||||||||||||||||||||
  17 uGlnValIleTyrAspGlyProAlaIleThrGluValAlaLeuLeuGlyG 34

 101 AAGAATATGCCGGTATGCGCCCCTNGATGAAAGTCAAGGAAGGCGATGCC 150
     ||||||||:::|||||||||||||:::|||||||:::|||||||||:::|||
  34 luGluTyrValGlyMetArgProSerMetLysIleLysGluGlyGluAla 50

 151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGNATCCGGGCGTGGT 200
     ||||||||||||||||||||||||||||||||||   ||||||||||||
  51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67

 201 GTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCATCCATCGCGGCGAAA 250
     ||||||||||||:::|||||||||||||||||||||||||||||||||||
  67 lPheThrAlaProAlaSerGlyLysIleAlaAlaIleHisArgGlyGluL 84

 251 AGCGCGTACTTCAGTCGGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  84 ysArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluIle 100

 301 GAGTTCGAACGCTACGCGCCCGAAGCGTTGGCAAACTTAAGCGGCGANGA 350
     ||||||||||||||:::|||||||||||||||:::|||||:::   ::
 101 GluPheGluArgTyrValProGluAlaLeuAlaLysLeuSerSerGluLy 117

 351 ANTNNGNNGCAATCTGATCCAATCCGGTTTGTGGACTGCGCTGCGTANCC 400
     :         ||||||||||||||||||||||||||||||||||||:::|
 117 sValArgArgAsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArgThrA 134

 401 GTCCGTTCAGCAAAATCCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTC 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150

 451 GTCAATGCGATGGACACCAATCCGCTNGCGGCAGACCCTGTGGTTGTGAT 500
     ||||||||||||||||||||||||||||||||||||:::|||:::||
 151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProThrValIleIl 167

 501 CAAAGAAGCCGNCGANGATTTCAGACGANGTNTGCTGGTATTGAGCCGTT 550
     ||||||||||:::   ||||||:::|||      ||||||||||||||||
```

```
 167 eLysGluAlaAlaGluAspPheLysArgGlyLeuLeuValLeuSerArgL 184

 551 TGACCGAGCGTAAAATCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCG 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 184 euThrGluArgLysIleHisValCysLysAlaAlaGlyAlaAspValPro 200

 601 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCC 650
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyGlyProHisPr 217

 651 GGCCGGTTTGAGTGGCACGCACATTCATTTCATTGAGCCGGTCGGTGCAA 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAlaA 234

 701 ACAAAACCGTTTGGACCATCAATTATCAAGATGTAATTGCCATCGGACGT 750
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 234 snLysThrValTrpThrIleAsnTyrGlnAspValIleAlaIleGlyArg 250

 751 TTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTGATTGCTTTGGGTGG 800
     ||||||:::||||||||||||||||||||||||||:::||||||||||||
 251 LeuPheValThrGlyArgLeuAsnThrGluArgValValAlaLeuGlyGl 267

 801 TTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTTGGGTGCGAAAG 850
     |   ||||||||||||||||||||||||||||||||||||||||||||||
 267 yLeuGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284

 851 TATCGCAAATTACTGCGGGCGAATTGGTTGACGCAGACAACCGCGTGATT 900
     ||||||||:::||||||||||||||||||||||||||||||||||||||||
 284 alSerGlnLeuThrAlaGlyGluLeuValAspAlaAspAsnArgValIle 300

 901 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTT 950
     ||||||||||||||||||||||||||:::||||||||||||||||||||||
 301 SerGlySerValLeuAsnGlyAlaIleAlaGlnGlyAlaHisAspTyrLe 317

 951 GGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAG 1000
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 317 uGlyArgTyrHisAsnGlnIleSerValIleGluGluGlyArgSerLysG 334

1001 AGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGT 1050
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 334 luLeuPheGlyTrpValAlaProGlnProAspLysTyrSerIleThrArg 350

1051 ACGACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTTCACGACAGC 1100
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 351 ThrThrLeuGlyHisPheLeuLysAsnLysLeuPheLysPheThrThrAl 367

1101 CGTCAACGGTGGCGACCGCGCCATGGTGCCGATTGGTACTTACGAGCGCG 1150
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 367 aValAsnGlyGlyAspArgAlaMetValProIleGlyThrTyrGluArgV 384

1151 TAATGCCGCTAGACATCCTGCCTACCCTGCTTTTGCGCGATTTAATCGTC 1200
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 384 alMetProLeuAspIleLeuProThrLeuLeuLeuArgAspLeuIleVal 400

1201 GGCGATACCGACAGCGCGCAAGCATTGGGTTGCTTGGAATTGGACGAAGA 1250
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 401 GlyAspThrAspSerAlaGlnAlaLeuGlyCysLeuGluLeuAspGluGl 417

1251 AGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATANGGCC 1300
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 417 uAspLeuAlaLeuCysSerPheValCysProGlyLysTyrGluTyrGlyP 434

1301 CGCTGTTGCGTAAGGTGCTGGAAACCNTTGAGAAGGAAGGC 1341
     ||||||||||||||||||||||||||||||||||||||||||
 434 roLeuLeuArgLysValLeuGluThrIleGluLysGluGly 447

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38563

seq_documentation_block:
ID   AAY38563 standard; Protein; 322 AA.
XX
AC   AAY38563;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria gonorrhoeae antigen encoded by a partial ORF22.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 124-125; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   322 AA;


alignment_scores:
           Quality: 1511.00              Length:    322
             Ratio:  4.827                Gaps:      0
 Percent Similarity:  97.205   Percent Identity: 91.304

alignment_block:
 US-09-303-518D-127 x AAY38563   ..

 Align seg 1/1  to: AAY38563  from: 1  to: 322

       1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGA 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProGl 17

      51 GCAAGTCATTTATGACGGGCCCGTCATTACCGAAGTCGCGTTGCTTGGCG 100
         ||||||||||||||||||||||:::|||||||||||||||||||||||||
      17 uGlnValIleTyrAspGlyProAlaIleThrGluValAlaLeuLeuGlyG 34

     101 AAGAATATGCCGGTATGCGCCCCTNGATGAAAGTCAAGGAAGGCGATGCC 150
         ||||||||:::|||||||||||||:::|||||||:::||||||||||:::|||
      34 luGluTyrValGlyMetArgProSerMetLysIleLysGluGlyGluAla 50

     151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGNATCCGGGCGTGGT 200
         ||||||||||||||||||||||||||||||||||||   ||||||||||||
      51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67

     201 GTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCATCCATCGCGGCGAAA 250
```

```
     ||||||||||||||:::|||||||||||||||||||||||||||||||||||
  67 lPheThrAlaProAlaSerGlyLysIleAlaAlaIleHisArgGlyGluL 84

 251 AGCGCGTACTTCAGTCGGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
  84 ysArgValLeuGlnSerValValIleAlaValGluGlyAsnAspGluIle 100

 301 GAGTTCGAACGCTACGCGCCCGAAGCGTTGGCAAACTTAAGCGGCGANGA 350
     ||||||||||||||:::|||||||||||||||:::||||||:::   ::
 101 GluPheGluArgTyrValProGluAlaLeuAlaLysLeuSerSerGluLy 117

 351 ANTNNGNNGCAATCTGATCCAATCCGGTTTGTGGACTGCGCTGCGTANCC 400
     :         |||||||||||||||||||||||||||||||||||||:::|
 117 sValArgArgAsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArgThrA 134

 401 GTCCGTTCAGCAAAATCCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTC 450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150

 451 GTCAATGCGATGGACACCAATCCGCTNGCGGCAGACCCTGTGGTTGTGAT 500
     ||||||||||||||||||||||||||||||||||||||:::|||:::||
 151 ValAsnAlaMetAspThrAsnProLeuAlaAlaAspProThrValIleIl 167

 501 CAAAGAAGCCGNCGANGATTTCAGACGANGTNTGCTGGTATTGAGCCGTT 550
     ||||||||||:::   ||||||:::|||      |||||||||||||||
 167 eLysGluAlaAlaGluAspPheLysArgGlyLeuLeuValLeuSerArgL 184

 551 TGACCGAGCGTAAAATCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCG 600
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 184 euThrGluArgLysIleHisValCysLysAlaAlaGlyAlaAspValPro 200

 601 TCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCC 650
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 201 SerGluAsnAlaAlaAsnIleGluThrHisGluPheGlyGlyProHisPr 217

 651 GGCCGGTTTGAGTGGCACGCACATTCATTTCATTGAGCCGGTCGGTGCAA 700
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 217 oAlaGlyLeuSerGlyThrHisIleHisPheIleGluProValGlyAlaA 234

 701 ACAAAACCGTTTGGACCATCAATTATCAAGATGTAATTGCCATCGGACGT 750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 234 snLysThrValTrpThrIleAsnTyrGlnAspValIleAlaIleGlyArg 250

 751 TTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTGATTGCTTTGGGTGG 800
     ||||||:::|||||||||||||||||||||||||||::||||||||||||
 251 LeuPheValThrGlyArgLeuAsnThrGluArgValValAlaLeuGlyGl 267

 801 TTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTTGGGTGCGAAAG 850
     |   ||||||||||||||||||||||||||||||||||||||||||||||
 267 yLeuGlnValAsnLysProArgLeuLeuArgThrValLeuGlyAlaLysV 284

 851 TATCGCAAATTACTGCGGGCGAATTGGTTGACGCAGACAACCGCGTGATT 900
     ||||||||:::|||||||||||||||||||||||||||||||||||||||
 284 alSerGlnLeuThrAlaGlyGluLeuValAspAlaAspAsnArgValIle 300

 901 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTT 950
     ||||||||||||||||||||||||||:::|||||||||||||||||||||
 301 SerGlySerValLeuAsnGlyAlaIleAlaGlnGlyAlaHisAspTyrLe 317

 951 GGGACGCTACCACAAT 966
     ||||||||||||||||
 317 uGlyArgTyrHisAsn 322
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY38560

seq_documentation_block:

```
ID   AAY38560 standard; Protein; 158 AA.
XX
AC   AAY38560;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis antigen encoded by a partial ORF22.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12025.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 122; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   158 AA;
```

alignment_scores:

Quality: 739.00 Length: 158
Ratio: 4.862 Gaps: 0
Percent Similarity: 96.203 Percent Identity: 92.405

alignment_block:
US-09-303-518D-127 x AAY38560 ..

Align seg 1/1 to: AAY38560 from: 1 to: 158

```
   1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGA 50
     ||||||||||||||||||||||||||||||||||||||||||||||||||
   1 MetIleLysIleLysLysGlyLeuAsnLeuProIleAlaGlyArgProGl 17

  51 GCAAGTCATTTATGACGGGCCCGTCATTACCGAAGTCGCGTTGCTTGGCG 100
     ||||::::::|||||||||||||:::|||||||||||||||||||||||||
  17 uGlnAlaValTyrAspGlyProAlaIleThrGluValAlaLeuLeuGlyG 34

 101 AAGAATATGCCGGTATGCGCCCCTNGATGAAAGTCAAGGAAGGCGATGCC 150
     |||||||||||||||||||||||:::|||||||||||||||||||||||||
  34 luGluTyrAlaGlyMetArgProSerMetLysValLysGluGlyAspAla 50

 151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGNATCCGGGCGTGGT 200
     ||||||||||||||||||||||||||||||||||   ||||||||||||
  51 ValLysLysGlyGlnValLeuPheGluAspLysLysAsnProGlyValVa 67

 201 GTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCATCCATCGCGGCGAAA 250
     ||||||||||||:::|||||||||||||||||||||||||||||||||||
  67 lPheThrAlaProAlaSerGlyLysIleAlaAlaIleHisArgGlyGluL 84
```

```
     251 AGCGCGTACTTCAGTCGGTCGTGATTGCCGTTGAAGGCAACGACGAAATC 300
         ||||||||||||||||||||||||||||||||||   ||||||||||||
      84 ysArgValLeuGlnSerValValIleAlaValGlu***AsnAspGluIle 100

     301 GAGTTCGAACGCTACGCGCCCGAAGCGTTGGCAAACTTAAGCGGCGANGA 350
         ||||||||||||||||||||||||||||||||||||||||||||||||   ||
     101 GluPheGluArgTyrAlaProGluAlaLeuAlaAsnLeuSerGlyGluGl 117

     351 ANTNNGNNGCAATCTGATCCAATCCGGTTTGTGGACTGCGCTGCGTANCC 400
         |         ||||||||||||||||||||||||||||||||||||||:::|
     117 uValArgArgAsnLeuIleGlnSerGlyLeuTrpThrAlaLeuArgThrA 134

     401 GTCCGTTCAGCAAAATCCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTC 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 rgProPheSerLysIleProAlaValAspAlaGluProPheAlaIlePhe 150

     451 GTCAATGCGATGGACACCAATCCG 474
         |||||||||||||||||||||||||
     151 ValAsnAlaMetAspThrAsnPro 158
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34439

seq_documentation_block:

ID AAY34439 standard; Protein; 451 AA.
XX
AC AAY34439;
XX
DT 25-AUG-1999 (first entry)
XX
DE Porphorymonas gingivalis protein PG1.
XX
KW Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
KW vaccine; antigenic.
XX
OS Porphorymonas gingivalis.
XX
PN WO9929870-A1.
XX
PD 17-JUN-1999.
XX
PF 10-DEC-1998; 98WO-AU01023.
XX
PR 04-AUG-1998; 98AU-0005028.
PR 10-DEC-1997; 97AU-0000839.
PR 31-DEC-1997; 97AU-0001182.
PR 30-JAN-1998; 98AU-0001546.
PR 10-MAR-1998; 98AU-0002264.
PR 09-APR-1998; 98AU-0002911.
PR 23-APR-1998; 98AU-0003128.
PR 05-MAY-1998; 98AU-0003338.
PR 22-MAY-1998; 98AU-0003654.
PR 29-JUL-1998; 98AU-0004917.
XX
PA (CSLC-) CSL LTD.
XX
PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX
DR WPI; 1999-385613/32.
DR N-PSDB; AAX91657.
XX
PT Antigenic Porphorymonas gingivalis peptides for preventing
PT gingivitis
XX
PS Claim 1; Page 417-418; 588pp; English.
XX
CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphorymonas gingivalis. Probes can
CC be used to detect Porphorymonas gingivalis in standard hybridisation
CC assays. Porphorymonas gingivalis is involved in periodontal disease
CC especially gingivitis.
XX
SQ Sequence 451 AA;

alignment_scores:

```
           Quality:  636.00              Length:    452
             Ratio:  2.208                 Gaps:      7
 Percent Similarity:  63.717   Percent Identity: 34.071
```

alignment_block:

US-09-303-518D-127 x AAY34439 ..

Align seg 1/1 to: AAY34439 from: 1 to: 451

```
       1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCG.. 48
         :::|||||||   ||||||||||||   |||   :::   |||:::|||
       4 ValIleLysThrLysLysGlyLeuAlaLeuAsnLeuLysGlyLysProLe 20

      49 .GAGCAAGTCATTTATGACGGGCCCGTCATTACCGAAGTCGCGTTGCTTG 97
          :::::::::   :::                        |||::::::
      20 uProGluMetLeuAlaGluProAlaGlnSerProThrTyrAlaValValP 37

      98 GCGAAGAATATGCCGGTATGCGCCCCTNGATGAAAGTCAAGGAAGGCGAT 147
         :::::::::   |||:::   |||   :::   ::::::   ||||||
      37 roAspAspPheGluGlyValIleProLysValThrAlaArgProGlyAsp 53

     148 GCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGNATCCGGGCGT 197
         |||:::   |||::::::|||:::::   |||      |||   ::
      54 LysValArgAlaGlySerAlaLeuMetHisHisLysAlaTyrProGluMe 70

     198 GGTGTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCATCCATCGCGGCG 247
         :   ||||||:::|||||||||||:::::   |||:::::|||||
      70 tLysPheThrSerProValSerGlyGluValIleAlaValAsnArgGlyA 87

     248 AAAAGCGCGTACTTCAGTCGGTCGTGATTGCCGTTGAAGGCAACGACGAA 297
         ||||||   :::   |||:::   :::      :::|||   :::|||
      87 laLysArgLysValLeuSerIleGluValLysProAspGlyLeuAsnGlu 103

     298 ATCGAG...TTCGAACGCTACGCGCCCGAAGCGTTGGCAAACTTAAGCGG 344
         |||   |||          |||:::|||          ||||||::
     104 TyrGluSerPheProValGlyAspProSerAla.....LeuSerAl 117

     345 CGANGAANTNNGNNGCAATCTGATCCAATCCGGTTTGTGGACTGCGCTGC 394
         :   :::           |||:::::|||||:::|||      ::::
     117 aGluGlnIleLysGluLeuLeuLeuSerSerGlyMetTrpGlyPheIleL 134

     395 GTANCCGTCCGTTCAGCAAAATCCCTGCCGTCGATGCCGAGCCGTTCGCC 444
         ::   ||||||::::::   :::   :::   |||      |||
     134 ysGlnArgProTyrAspIleValAlaThrProAspIleAlaProArgAsp 150

     445 ATCTTCGTCAATGCGATGGACACCAATCCGCTNGCGGCAGACCCTGTGGT 494
         |||::::::::::|||      |||   |||||||||   |||
     151 IleTyrIleThrAlaAsnPheThrAlaProLeuAlaProAspPheAspPh 167

     495 TGTGATCAAAGAAGCCGNCGANGATTTCAGACGANGTNTGCTGGTATTGA 544
          :::::::::                 ::::::   :::      :::|||:
     167 eIleValArgGlyGluGluArgAlaLeuGlnThrAlaIleAspAlaLeuA 184

     545 GCCGTTTGACCGAGCGTAAAATCCATGTGTGTAAGGCAGCTGGCGCAGAC 594
         :::::||||||      |||:::::::|||            |||::::::
     184 laLysLeuThrThrGlyLysValTyrValGlyLeuLysProGlySerSer 200

     595 GTGCCGTCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCC 644
         :::      :::||||||      :::|||:::|||            |||||
     201 LeuGlyLeuHisAsnAlaGluIleValGluValHis.....GlyPr 214

     645 GCATCCGGCCGGTTTGAGTGGCACGCACATTCATTTCATTGAGCCGGTCG 694
         ||||||||||||      |||:::   |||:::      :::|||::::
     214 oHisProAlaGlyAsnValGlyValLeuIleAsnHisThrLysProIleA 231
```

```
 695 GTGCAAACAAAACCGTTTGGACCATCAATTATCAAGATGTAATTGCCATC 744
     ::   ::::::|||||||||||||:::::::      |||:::|||:::|||
 231 snArgGlyGluThrValTrpThrLeuLysAlaThrAspLeuIleValIle 247

 745 GGACGTTTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTGATTGCTTT 794
     ||||||::::::   ||||||:::   :::      |||::|||||||::
 248 GlyArgPheLeuLeuThrGlyLysAlaAspPheThrArgMetIleAlaMe 264

 795 GGGTGGTTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTTGGGTG 844
     :   ||||||::::::               :::|||   :::   |||:
 264 tThrGlySerAspAlaAlaAlaHisGlyTyrValArgIleMetProGlyC 281

 845 CGAAAGTATCGCAAATTACTGCGGGCGAATTGGTT......GACGCAGAC 888
     :::::|||              |||:::|||:::      ::::::
 281 ysAsnValPheAlaSerPheProGlyArgLeuThrIleLysGluSerHis 297

 889 AACCGCGTGATTTCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGC 938
     :::||||||||:::|||:::||||||:::|||
 298 GluArgValIleAspGlyAsnValLeuThrGlyLysLysLeuCysGluLy 314

 939 GCACGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAG 988
      :::   :::|||:::         :::||||||:::||||||   ||||
 314 sGluProPheLeuSerAlaArgCysAspGlnIleThrValIleProGluG 331

 989 GCCGCAGC...AAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAA 1035
     ||   :::      ||||||||||||||:::||||||:::   |||:::
 331 lyAspAspValAspGluLeuPheGlyTrpAlaAlaProArgLeuAspGln 347

1036 TACTCCATCACGCGTACGACCCTCGGCCATTTCCTG...AAAAACAAACT 1082
     ||||||:::::|||:::   ::::::   :::      |||||||||
 348 TyrSerMetSerArgAlaTyrPheSerTrpLeuGlnGlyLysAsnLysGl 364

1083 CTTCAAGTTCACGACAGCCGTCAACGGTGGCGACCGCGCCATGGTGCCGA 1132
      :::   :::   :::   ::::::||||||:::|||||||||:::
 364 uTyrValLeuAspAlaArgIleLysGlyGlyGluArgAlaMetIleMetS 381

1133 TTGGTACTTACGAGCGCGTAATGCCGCTAGACATCCTGCCTACCCTGCTT 1182
       :::   |||:::||||||:::|||:::||||||   |||      |||
 381 erAsnGluTyrAspArgValPheProMetAspIleTyrProGluTyrLeu 397

1183 TTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAAGCATTGGGTTG 1232
     |||:::   :::|||:::   |||   |||:::   :::   ||||||
 398 LeuLysAlaIleIleAlaPheAspIleAspLysMetGluAspLeuGlyIl 414

1233 CTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGG 1282
        |||:::      ||||||:::|||   |||:::||||||     :
 414 eTyrGluValAlaProGluAspPheAlaThrCysGluPheValAspThrS 431

1283 GCAAATACGAATANGGCCCGCTGTTGCGTAAGGTGCTGGAAACCNTTGAG 1332
     ::|||   |||         ::::::|||:::   |||:::
 431 erLysIleGluLeuGlnArgIleValArgGluGlyLeuAspMetLeuTyr 447

1333 AAGGAA 1338
     ||||||
 448 LysGlu 449
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34318

seq_documentation_block:

```
ID   AAY34318 standard; Protein; 454 AA.
XX
AC   AAY34318;
XX
DT   25-AUG-1999  (first entry)
XX
DE   Porphorymonas gingivalis protein PG1.
XX
KW   Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
KW   vaccine; antigenic.
XX
OS   Porphorymonas gingivalis.
XX
PN   WO9929870-A1.
XX
PD   17-JUN-1999.
XX
PF   10-DEC-1998;   98WO-AU01023.
XX
PR   04-AUG-1998;   98AU-0005028.
PR   10-DEC-1997;   97AU-0000839.
PR   31-DEC-1997;   97AU-0001182.
PR   30-JAN-1998;   98AU-0001546.
PR   10-MAR-1998;   98AU-0002264.
PR   09-APR-1998;   98AU-0002911.
PR   23-APR-1998;   98AU-0003128.
PR   05-MAY-1998;   98AU-0003338.
PR   22-MAY-1998;   98AU-0003654.
PR   29-JUL-1998;   98AU-0004917.
XX
PA   (CSLC-) CSL LTD.
XX
PI   Agius CT,  Barr IG,  Hocking DM,  Margetts MB,  Patterson MA;
PI   Ross BC,  Rothel LJ,  Webb EA;
XX
DR   WPI; 1999-385613/32.
DR   N-PSDB; AAX91536.
XX
PT   Antigenic Porphorymonas gingivalis peptides for preventing
PT   gingivitis
XX
PS   Claim 1; Page 277; 588pp; English.
XX
CC   AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC   Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC   AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC   isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC   activity with a vaccine mechanism of action. The PG polypeptides can be
CC   used as vaccines especially against Porphorymonas gingivalis. Probes can
CC   be used to detect Porphorymonas gingivalis in standard hybridisation
CC   assays. Porphorymonas gingivalis is involved in periodontal disease
CC   especially gingivitis.
XX
SQ   Sequence   454 AA;
```

alignment_scores:

```
            Quality:  636.00              Length:    452
              Ratio:  2.208                 Gaps:      7
 Percent Similarity:  63.717   Percent Identity: 34.071
```

alignment_block:

US-09-303-518D-127 x AAY34318 ..

Align seg 1/1 to: AAY34318 from: 1 to: 454

```
   1 ATGATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCG.. 48
     :::||||||   ||||||||||||   |||   :::   |||:::|||
   7 ValIleLysThrLysLysGlyLeuAlaLeuAsnLeuLysGlyLysProLe 23

  49 .GAGCAAGTCATTTATGACGGGCCCGTCATTACCGAAGTCGCGTTGCTTG 97
        :::::::::   :::                     |||::::::
  23 uProGluMetLeuAlaGluProAlaGlnSerProThrTyrAlaValValP 40

  98 GCGAAGAATATGCCGGTATGCGCCCCTNGATGAAAGTCAAGGAAGGCGAT 147
     :::::::::   |||:::   |||   :::   ::::::   ||||||
  40 roAspAspPheGluGlyValIleProLysValThrAlaArgProGlyAsp 56

 148 GCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGNATCCGGGCGT 197
       |||:::   |||::::::|||::::::   |||      |||   ::
  57 LysValArgAlaGlySerAlaLeuMetHisHisLysAlaTyrProGluMe 73

 198 GGTGTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCATCCATCGCGGCG 247
     :   ||||||:::|||||||||||:::::   |||:::::|||||
  73 tLysPheThrSerProValSerGlyGluValIleAlaValAsnArgGlyA 90

 248 AAAAGCGCGTACTTCAGTCGGTCGTGATTGCCGTTGAAGGCAACGACGAA 297
```

```
 90 laLysArgLysValLeuSerIleGluValLysProAspGlyLeuAsnGlu 106

298 ATCGAG...TTCGAACGCTACGCGCCCGAAGCGTTGGCAAACTTAAGCGG 344
107 TyrGluSerPheProValGlyAspProSerAla.....LeuSerAl 120

345 CGANGAANTNNGNNGCAATCTGATCCAATCCGGTTTGTGGACTGCGCTGC 394
120 aGluGlnIleLysGluLeuLeuLeuSerSerGlyMetTrpGlyPheIleL 137

395 GTANCCGTCCGTTCAGCAAAATCCCTGCCGTCGATGCCGAGCCGTTCGCC 444
137 ysGlnArgProTyrAspIleValAlaThrProAspIleAlaProArgAsp 153

445 ATCTTCGTCAATGCGATGGACACCAATCCGCTNGCGGCAGACCCTGTGGT 494
154 IleTyrIleThrAlaAsnPheThrAlaProLeuAlaProAspPheAspPh 170

495 TGTGATCAAAGAAGCCGNCGANGATTTCAGACGANGTNTGCTGGTATTGA 544
170 eIleValArgGlyGluGluArgAlaLeuGlnThrAlaIleAspAlaLeuA 187

545 GCCGTTTGACCGAGCGTAAAATCCATGTGTGTAAGGCAGCTGGCGCAGAC 594
187 laLysLeuThrThrGlyLysValTyrValGlyLeuLysProGlySerSer 203

595 GTGCCGTCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCC 644
204 LeuGlyLeuHisAsnAlaGluIleValGluValHis.....GlyPr 217

645 GCATCCGGCCGGTTTGAGTGGCACGCACATTCATTTCATTGAGCCGGTCG 694
217 oHisProAlaGlyAsnValGlyValLeuIleAsnHisThrLysProIleA 234

695 GTGCAAACAAAACCGTTTGGACCATCAATTATCAAGATGTAATTGCCATC 744
234 snArgGlyGluThrValTrpThrLeuLysAlaThrAspLeuIleValIle 250

745 GGACGTTTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTGATTGCTTT 794
251 GlyArgPheLeuLeuThrGlyLysAlaAspPheThrArgMetIleAlaMe 267

795 GGGTGGTTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTTGGGTG 844
267 tThrGlySerAspAlaAlaAlaHisGlyTyrValArgIleMetProGlyC 284

845 CGAAAGTATCGCAAATTACTGCGGGCGAATTGGTT......GACGCAGAC 888
284 ysAsnValPheAlaSerPheProGlyArgLeuThrIleLysGluSerHis 300

889 AACCGCGTGATTTCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGC 938
301 GluArgValIleAspGlyAsnValLeuThrGlyLysLysLeuCysGluLy 317

939 GCACGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAG 988
317 sGluProPheLeuSerAlaArgCysAspGlnIleThrValIleProGluG 334

989 GCCGCAGC...AAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAA 1035
334 lyAspAspValAspGluLeuPheGlyTrpAlaAlaProArgLeuAspGln 350

1036 TACTCCATCACGCGTACGACCCTCGGCCATTTCCTG...AAAAACAAACT 1082
351 TyrSerMetSerArgAlaTyrPheSerTrpLeuGlnGlyLysAsnLysGl 367

1083 CTTCAAGTTCACGACAGCCGTCAACGGTGGCGACCGCGCCATGGTGCCGA 1132
367 uTyrValLeuAspAlaArgIleLysGlyGlyGluArgAlaMetIleMetS 384

1133 TTGGTACTTACGAGCGCGTAATGCCGCTAGACATCCTGCCTACCCTGCTT 1182
384 erAsnGluTyrAspArgValPheProMetAspIleTyrProGluTyrLeu 400

1183 TTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAAGCATTGGGTTG 1232
401 LeuLysAlaIleIleAlaPheAspIleAspLysMetGluAspLeuGlyIl 417

1233 CTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGG 1282
417 eTyrGluValAlaProGluAspPheAlaThrCysGluPheValAspThrS 434

1283 GCAAATACGAATANGGCCCGCTGTTGCGTAAGGTGCTGGAAACCNTTGAG 1332
434 erLysIleGluLeuGlnArgIleValArgGluGlyLeuAspMetLeuTyr 450

1333 AAGGAA 1338
451 LysGlu 452
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75273

seq_documentation_block:

ID AAY75273 standard; Protein; 120 AA.
XX
AC AAY75273;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 628 protein sequence SEQ ID NO:2020.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR N-PSDB; AAZ54035.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2; Page 1004; 1453pp; English.
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to

```
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   120 AA;


alignment_scores:
           Quality:  563.00            Length:    120
             Ratio:  4.896              Gaps:      0
 Percent Similarity:  95.833   Percent Identity: 94.167

alignment_block:
 US-09-303-518D-127/rev x AAY75273   ..

 Align seg 1/1  to: AAY75273  from: 1  to: 120

     674 ATGTGCGTGCCACTCAAACCGGCCGGATGCGGGCCGCCGAATTCATGTGT 625
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetCysValProLeuLysProAlaGlyCysGlyProProAsnSerCysVa 17

     624 TTCGATGTTGGCAGCATTTTCAGACGGCACGTCTGCGCCAGCTGCCTTAC 575
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 lSerMetLeuAlaAlaPheSerAspGlyThrSerAlaProAlaAlaLeuH 34

     574 ACACATGGATTTTACGCTCGGTCAAACGGCTCAATACCAGCANACNTCGT 525
         ||||||||||||||||||||||||||||||||||||||      |||
      34 isThrTrpIleLeuArgSerValLysArgLeuAsnThrSerLysProArg 50

     524 CTGAAATCNTCGNCGGCTTCTTTGATCACAACCACAGGGTCTGCCGCNAG 475
         ||||||||||||:::|||||||||||||||||||||||||||||||||||
      51 LeuLysSerSerAlaAlaSerLeuIleThrThrThrGlySerAlaAlaSe 67

     474 CGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCGA 425
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 rGlyLeuValSerIleAlaLeuThrLysMetAlaAsnGlySerAlaSerT 84

     424 CGGCAGGGATTTTGCTGAACGGACGGNTACGCAGCGCAGTCCACAAACCG 375
         |||||||||||||||||||||||||   |||||||||||||||||||||
      84 hrAlaGlyIleLeuLeuAsnGlyArgValArgSerAlaValHisLysPro 100

     374 GATTGGATCAGATTGCNNCNNANTTCNTCGCCGCTTAAGTTTGCCAACGC 325
         ||||||||||||||      :::||||||||||||||||||||||||||
     101 AspTrpIleArgLeuArgArgThrSerSerProLeuLysPheAlaAsnAl 117

     324 TTCGGGCGCG 315
         ||||||||||
     117 aSerGlyAla 120

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75272

seq_documentation_block:
ID   AAY75272 standard; Protein; 120 AA.
XX
AC   AAY75272;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 628 protein sequence SEQ ID NO:2018.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ54034.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 1003; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAZ75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   120 AA;


alignment_scores:
           Quality:  536.00            Length:    120
             Ratio:  4.702              Gaps:      0
 Percent Similarity:  95.000   Percent Identity: 90.000

alignment_block:
 US-09-303-518D-127/rev x AAY75272   ..

 Align seg 1/1  to: AAY75272  from: 1  to: 120

     674 ATGTGCGTGCCACTCAAACCGGCCGGATGCGGGCCGCCGAATTCATGTGT 625
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetCysValProLeuLysProAlaGlyCysGlyProProAsnSerCysVa 17

     624 TTCGATGTTGGCAGCATTTTCAGACGGCACGTCTGCGCCAGCTGCCTTAC 575
         ||||||||||||||||||||||||||||||||||||||||||||||||||:
      17 lSerMetLeuAlaAlaPheSerAspGlyThrSerAlaProAlaAlaLeuG 34

     574 ACACATGGATTTTACGCTCGGTCAAACGGCTCAATACCAGCANACNTCGT 525
         ::|||||||||||||||||||||||||||||||||||:::      |||
      34 lnThrTrpIleLeuArgSerValLysArgLeuAsnThrAsnArgProArg 50

     524 CTGAAATCNTCGNCGGCTTCTTTGATCACAACCACAGGGTCTGCCGCNAG 475
         ||||||||||||:::||||||||||||   |||:::|||||||||||||||
      51 LeuLysSerSerAlaAlaSerLeuIleMetThrValGlySerAlaAlaSe 67

     474 CGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCGA 425
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 rGlyLeuValSerIleAlaLeuThrLysMetAlaAsnGlySerAlaSerT 84

     424 CGGCAGGGATTTTGCTGAACGGACGGNTACGCAGCGCAGTCCACAAACCG 375
```

```
      |||||||||||||||||||||||||   ||||||||||||||||||||
   84 hrAlaGlyIleLeuLeuAsnGlyArgValArgSerAlaValHisLysPro 100

  374 GATTGGATCAGATTGCNNCNNANTTCNTCGCCGCTTAAGTTTGCCAACGC 325
      ||||||||||||||||      :::|||||||||||||||||||||:::||
  101 AspTrpIleArgLeuArgArgThrSerSerProLeuLysPheAlaSerAl 117

  324 TTCGGGCGCG 315
      ||||||||||
  117 aSerGlyAla 120

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75271

seq_documentation_block:
ID   AAY75271 standard; Protein; 119 AA.
XX
AC   AAY75271;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 628 protein sequence SEQ ID NO:2016.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ54033.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 1003; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   119 AA;
```

```
alignment_scores:
            Quality:  487.50              Length:    119
              Ratio:  4.432                 Gaps:      1
 Percent Similarity:  92.437   Percent Identity: 84.874

alignment_block:
 US-09-303-518D-127/rev x AAY75271   ..

 Align seg 1/1  to: AAY75271  from: 1  to: 119

  674 ATGTGCGTGCCACTCAAACCGGCCGGATGCGGGCCGCCGAATTCATGTGT 625
      ||||||||||||||||||||||||||||||||||||||||||||||||||
    1 MetCysValProLeuLysProAlaGlyCysGlyProProAsnSerCysVa 17

  624 TTCGATGTTGGCAGCATTTTCAGACGGCACGTCTGCGCCAGCTGCCTTAC 575
      ||||:::|||||||||||||||||||||||||||||||||||||||||||
   17 lSerIleLeuAlaAlaPheSerAspGlyThrSerAlaProAlaAlaLeuH 34

  574 ACACATGGATTTTACGCTCGGTCAAACGGCTCAATACCAGCANACNTCGT 525
      ||||||||||||||||||||||:::|||||||||||||:::      |||
   34 isThrTrpIleLeuArgSerValArgArgLeuAsnThrAsnArgProArg 50

  524 CTGAAATCNTCGNCGGCTTCTTTGATCACAACCACAGGGTCTGCCGCNAG 475
      |||||||||||||:::|||||||||:::   |||:::|||||||||||||
   51 LeuLysSerSerAlaAlaSerLeuMetMetThrValGlySerAlaAlaSe 67

  474 CGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCGA 425
      ||||||||||||||||||||||||||||||||||||||||||||||||||
   67 rGlyLeuValSerIleAlaLeuThrLysMetAlaAsnGlySerAlaSerT 84

  424 CGGCAGGGATTTTGCTGAACGGACGGNTACGCAGCGCAGTCCACAAACCG 375
      ||||||||||||||||||||||||||   ||||||||||||||||||||
   84 hrAlaGlyIleLeuLeuAsnGlyArgValArgSerAlaValHisLysPro 100

  374 GATTGGATCAGATTGCNNCNNANTTCNTCGCCGCTTAAGTTTGCCAACGC 325
      |||   |||||||||      :::   |||   |||:::||||||:::||
  101 Asp...IleArgLeuArgArgThrPheSerLeuLeuAsnPheAlaSerAl 116

  324 TTCGGGC 318
      |||||||
  116 aSerGly 118

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY82082

seq_documentation_block:
ID   AAY82082 standard; Protein; 467 AA.
XX
AC   AAY82082;
XX
DT   01-JUN-2000  (first entry)
XX
DE   Chlamydia pneumoniae antigen CPN100605 protein SEQ ID NO:2.
XX
KW   Chlamydia pneumoniae; antigen; CPN100605 protein; immunisation;
KW   vaccine; infection; antibacterial; antiinflammatory; bronchitis;
KW   community acquired pneumonia; upper respiratory tract infection;
KW   sinusitis.
XX
OS   Chlamydia pneumoniae.
XX
PN   WO200006742-A2.
XX
PD   10-FEB-2000.
XX
PF   27-JUL-1999;   99WO-IB01331.
XX
PR   27-JUL-1998;   98US-0094195.
PR   26-JUL-1999;   99US-0361443.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
```

PI Murdin AD, Oomen RP;
XX
DR WPI; 2000-205466/18.
DR N-PSDB; AAZ95378.
XX
PT Chlamydia pneumoniae antigens used for immunization and protection
PT against Chlamydia diseases -
XX
PS Claim 6; Fig 1; 48pp; English.
XX
CC The present sequence represents the Chlamydia pneumoniae antigen
CC CPN100605 protein. The CPN100605 protein has antibacterial and
CC antiinflammatory activities. The Chlamydia pneumoniae CPN100605
CC polynucleotide and protein can be used in vaccination methods for
CC preventing and treating Chlamydia infection (e.g. infections caused by
CC C. trachomatis, C. psittaci, C. pneumoniae or C. pecorum). The
CC polynucleotide can be used to produce the protein recombinantly, in the
CC construction of vaccine vectors, as a vaccine agent, and in the
CC construction of an attenuated Chlamydia strain. The protein are also be
CC useful as a vaccine agent, and for the preparation of medicaments for
CC treating or preventing Chlamydia infection, e.g. community acquired
CC pneumonia, and upper respiratory tract infections such as bronchitis and
CC sinusitis.
XX
SQ Sequence 467 AA;

alignment_scores:
Quality: 447.50 Length: 468
Ratio: 1.565 Gaps: 15
Percent Similarity: 61.111 Percent Identity: 29.487

alignment_block:
US-09-303-518D-127 x AAY82082 ..

Align seg 1/1 to: AAY82082 from: 1 to: 467

```
  4 ATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCA 53
    |||   :::::::::|||||||:::|||   :::   |||   |||:::::
  3 IleThrValAsnArgGlyLeuAspLeuSerLeuGlnGlySerProLysGl 19

 54 AGTCATTTATGACGGGCCCGTCATTACCGAAGTCGCGTTGCTTGGCGAAG 103
    :   :::   :::   :::      |||   :::   :::
 19 uSerGlyPheTyrAsnLysIleAspProGluPheValSerIle..... 33

104 AATATGCCGGTATGCGCCCC.................TNGATGAAAGTC 135
              :::||||||                        :::||||||
 34 ........AspLeuArgProPheGlnProLeuSerLeuLysLeuLysVal 47

136 AAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAA 185
    ::::::||||||||||||   :::|||      :::   |||   |||
 48 GluGlnGlyAspAlaValCysSerGlyAlaProIleAlaGluTyrLysHi 64

186 GNATCCGGGCGTGGTGTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCA 235
        |||::::::   :::|||:::   |||||||||   ::::::||||
 64 sPheProAsnThrTyrIleThrSerHisValSerGlyValValThrAlaI 81

236 TCCATCGCGGCGAAAAGCGCGTACTTCAGTCGGTCGTGATT...GCCGTT 282
    ||:::||||||:::||||||   |||   :::|||:::|||      :::
 81 leArgArgGlyAsnLysArgSerLeuLeuAspValIleIleLysLysThr 97

283 GAAGGCAACGACGAAATCGAGTTCGAACGCTACGCGCCCGAAGCGTTGGC 332
       |||      :::   |||:::                      |||
 98 ProGlyProThrSerThrGluTyr.............ThrTyrAspLeuGl 110

333 AAACTTAAGCGGCGANGAANTNNGNNGCAATCTGATCCAATCCGGTTTGT 382
     :::||||||   ::::::   :::      :::   ::::::||||||:
110 nThrLeuSerArgSerAspLeuSerGluIlePheLysGluAsnGlyLeuP 127

383 GGACTGCGCTGCGTANCCGTCCGTTCAGCAAAATCCCTGCCGTCGATGCC 432
    :::::   ::::::   ||||||||||:::   ||||||||||:::   :::
127 heAlaLeuIleLysGlnArgProPheAsp...IleProAlaIleProThr 142

433 GAG...CCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTNGC 479
    :::   |||      :::|||:::|||      |||::::::|||:::::
143 GlnThrProArgAspValPheIleAsnLeuAlaAspAsnArgProPheTh 159

480 GGCAGACCCTGTGGTTGTGATC.................AAAGAAGCCG 511
    :   :::|||         :::                  :::|||:::
159 rProSerProGluLysHisLeuAlaLeuPheSerSerArgGluGluGlyP 176

512 NCGANGATTTCAGACGANGTNTGCTGGTATTGAGCCGTTTGACCGAGCGT 561
          |||                :::::::::::|||
176 heTyrValPheValValGlyValArgAlaIleAlaLysLeuPheGlyLeu 192

562 AAAATCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGC 611
    :::   |||:::                      :::|||:::::::::
193 ArgProHisIleValPheArgAspArgLeuThrLeuProThrGlnGluLe 209

612 TGCCAACATC...GAAACACATGAATTCGGCGGCCCGCATCCGGCCGGTT 658
       :::|||   :::   |||      :::||||||   |||:::|||
209 uLysThrIleAlaHisLeuHisThrValSerGlyProPheProSerGlyS 226

659 TGAGTGGCACGCACATTCATTTCATTGAGCCGGTCGGTGCAAACAAA... 705
        :::   |||||||||   :::   |||:::      :::|||
226 erProSerIleHisIleHisSerValAlaProIleThrAsnGluLysGlu 242

706 ACCGTTTGGACCATCAATTATCAAGATGTAATTGCCATCGGACGTTTGTT 755
    :::|||:::|||:::::::::|||||||||:::::|||||||:::|||||
243 ValValPheThrLeuSerPheGlnAspValLeuThrIleGlyHisLeuPh 259

756 TGCAACAGGCCGTCTGAACACCGAGCGCGTGATTGCTTTGGGTGGTTCTC 805
    |      ||||||:::      |||:::|||   ||||||:::|||:::
259 eLeuLysGlyArgIleLeuHisGluGlnValThrAlaLeuAlaGlyThrA 276

806 AAGTCAACAAACCA......CGCCTCTTGCGTACCGTTTTGGGTGCGAAA 849
      :::::::::         |||   :::   |||:::   ||||||::
276 laLeuLysSerSerLeuArgArgTyrValIleThrThrLysGlyAlaSer 292

850 GTATCGCAA...ATTACTGCGGGCGAATTGGTTGACGCAGACAACCGCGT 896
      |||:::   |||:::   :::::::::   |||   |||:::   ::
293 PheSerSerLeuIleAsnLeuAsnAspIleSerAspAsnAspThr...Le 308

897 GATTTCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGAT. 945
    :|||||||||:::   |||:::|||   :::   :::      ::::::
308 uIleSerGlyAspProLeuThrGlyArgLeuCysLysLysGluGluGluP 325

946 ..TATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGC 993
      :::||||||         ::::::|||||||||:::::::::
325 roPheLeuGlyPheArgAspHisSerIleSerValLeuHisAsnProThr 341

994 AGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCAT 1043
    ::::::|||||||||:::::::::            :::|||   :::::
342 LysArgGluLeuPheSerPheLeuArgIleGlyPheAsnLysProThrPh 358

1044 CACGCGTACGACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTTCA 1093
     :|||:::|||   |||:::   |||:::|||:::|||         :::|
358 eThrLysThrTyrLeuSerGlyPhePheLysLysLysArg...ThrTyrT 374

1094 CG.........ACAGCCGTCAACGGTGGCGACCGCGCCATGGTGCCGATT 1134
     ||         |||   ::::::|||      |||   ::::::
374 hrAsnProAspThrAsnLeuHisGlyGluThrArgProIleIleAspThr 390

1135 GGTACTTACGAGCGCGTAATGCCGCTAGACATCCTGCCTACCCTGCTTTT 1184
         |||::::::|||||||||:::   |||      :::   |||::
391 AspIleTyrAspLysValMetProMetArgIleProValValProLeuIl 407

1185 GCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAAGCATTGGGTTGCT 1234
     ::::   :::|||:::   :::   |||   |||:::   ||||||   |
407 eLysAlaValIleThrLysAsnPheAspLeuAlaAsnGluLeuGlyPheL 424

1235 TGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGC 1284
     |||||:::      |||||||:::||||||   :::::::::   |||:::
424 euGluValCysGlyGluAspPheAlaLeuProThrLeuIleAspProSer 440

1285 AAATACGAATANGGCCCGCTGTTGCGTAAGGTGCTGGAAACCNTTGAGAA 1334
```

```
      |||   |||         ::::::::::::   |||           ||
  441 LysThrGluMetLeuThrIleValLysGluSerLeuIleGluTyrAlaLy 457

 1335 GGAA 1338
      ||||
  457 sGlu 458
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY35375

seq_documentation_block:

ID AAY35375 standard; Protein; 469 AA.
XX
AC AAY35375;
XX
DT 13-SEP-1999 (first entry)
XX
DE Chlamydia pneumoniae transmembrane protein sequence.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.
XX
OS Chlamydia pneumoniae.
XX
PN WO9927105-A2.
XX
PD 03-JUN-1999.
XX
PF 20-NOV-1998; 98WO-IB01890.
XX
PR 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX
PA (GEST) GENSET.
XX
PI Griffais R;
XX
DR WPI; 1999-357842/30.
XX
PT Genome sequence of Chlamydia pneumoniae
XX
PS Page 1170-1171; Disclosure; 1912pp; English.
XX
CC AAY34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotides sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX
SQ Sequence 469 AA;

```
alignment_scores:
            Quality:  447.50              Length:    468
              Ratio:  1.565                 Gaps:     15
 Percent Similarity:  61.111   Percent Identity: 29.487

alignment_block:
 US-09-303-518D-127 x AAY35375   ..

 Align seg 1/1  to: AAY35375  from: 1  to: 469

       4 ATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCA 53
         |||   :::::::::||||||:::|||   :::   |||   |||:::::
       5 IleThrValAsnArgGlyLeuAspLeuSerLeuGlnGlySerProLysGl 21

      54 AGTCATTTATGACGGGCCCGTCATTACCGAAGTCGCGTTGCTTGGCGAAG 103
         :      :::   :::   :::      |||   :::   :::
      21 uSerGlyPheTyrAsnLysIleAspProGluPheValSerIle..... 35

     104 AATATGCCGGTATGCGCCCC.................TNGATGAAAGTC 135
                 ::::||||||                   :::|||||||
      36 ........AspLeuArgProPheGlnProLeuSerLeuLysLeuLysVal 49

     136 AAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAA 185
         ::::::||||||||||||   :::|||      :::   |||   |||
      50 GluGlnGlyAspAlaValCysSerGlyAlaProIleAlaGluTyrLysHi 66

     186 GNATCCGGGCGTGGTGTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCA 235
           |||::::::   :::|||:::   |||||||||   :::::::||||
      66 sPheProAsnThrTyrIleThrSerHisValSerGlyValValThrAlaI 83

     236 TCCATCGCGGCGAAAAGCGCGTACTTCAGTCGGTCGTGATT...GCCGTT 282
         ||:::||||||:::|||||||   |||   :::|||:::|||      :::
      83 leArgArgGlyAsnLysArgSerLeuLeuAspValIleIleLysLysThr 99

     283 GAAGGCAACGACGAAATCGAGTTCGAACGCTACGCGCCCGAAGCGTTGGC 332
         |||      :::   |||:::                          |||
     100 ProGlyProThrSerThrGluTyr.....ThrTyrAspLeuGl 112

     333 AAACTTAAGCGGCGANGAANTNNGNNGCAATCTGATCCAATCCGGTTTGT 382
          :::|||||||   ::::::   :::       :::   ::::::||||||:
     112 nThrLeuSerArgSerAspLeuSerGluIlePheLysGluAsnGlyLeuP 129

     383 GGACTGCGCTGCGTANCCGTCCGTTCAGCAAAATCCCTGCCGTCGATGCC 432
         :::::   ::::::   ||||||||||:::   ||||||||||:::   :::
     129 heAlaLeuIleLysGlnArgProPheAsp...IleProAlaIleProThr 144

     433 GAG...CCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTNGC 479
         :::   |||      :::|||:::|||      |||::::::|||:::::
     145 GlnThrProArgAspValPheIleAsnLeuAlaAspAsnArgProPheTh 161

     480 GGCAGACCCTGTGGTTGTGATC.................AAAGAAGCCG 511
         :   :::|||         :::                  :::|||:::
     161 rProSerProGluLysHisLeuAlaLeuPheSerSerArgGluGluGlyP 178

     512 NCGANGATTTCAGACGANGTNTGCTGGTATTGAGCCGTTTGACCGAGCGT 561
               |||              ::::::::::::|||
     178 heTyrValPheValValGlyValArgAlaIleAlaLysLeuPheGlyLeu 194

     562 AAAATCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGC 611
         :::   |||:::                  :::|||:::::::::
     195 ArgProHisIleValPheArgAspArgLeuThrLeuProThrGlnGluLe 211

     612 TGCCAACATC...GAAACACATGAATTCGGCGGCCCGCATCCGGCCGGTT 658
             :::|||   :::   |||      :::|||||   |||:::|||
     211 uLysThrIleAlaHisLeuHisThrValSerGlyProPheProSerGlyS 228

     659 TGAGTGGCACGCACATTCATTTCATTGAGCCGGTCGGTGCAAACAAA... 705
             :::   |||||||||   :::   |||:::      :::|||
     228 erProSerIleHisIleHisSerValAlaProIleThrAsnGluLysGlu 244

     706 ACCGTTTGGACCATCAATTATCAAGATGTAATTGCCATCGGACGTTTGTT 755
         :::|||:::|||::::::::::|||||||||:::::::||||||:::|||||
     245 ValValPheThrLeuSerPheGlnAspValLeuThrIleGlyHisLeuPh 261

     756 TGCAACAGGCCGTCTGAACACCGAGCGCGTGATTGCTTTGGGTGGTTCTC 805
         |      |||||||:::      |||:::|||   |||||||:::|||:::
     261 eLeuLysGlyArgIleLeuHisGluGlnValThrAlaLeuAlaGlyThrA 278

     806 AAGTCAACAAACCA......CGCCTCTTGCGTACCGTTTTGGGTGCGAAA 849
           :::::::::         |||   :::   |||:::   |||||||:::
     278 laLeuLysSerSerLeuArgArgTyrValIleThrThrLysGlyAlaSer 294

     850 GTATCGCAA...ATTACTGCGGGCGAATTGGTTGACGCAGACAACCGCGT 896
             |||:::   |||:::   :::::::::   |||   |||:::   ::
     295 PheSerSerLeuIleAsnLeuAsnAspIleSerAspAsnAspThr...Le 310

     897 GATTTCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGAT. 945
         :|||||||||:::   |||:::|||   :::   :::      ::::::
```

```
 310 uIleSerGlyAspProLeuThrGlyArgLeuCysLysLysGluGluGluP 327

 946 ..TATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGC 993
       :::||||||         ::::::||||||||:::::::::
 327 roPheLeuGlyPheArgAspHisSerIleSerValLeuHisAsnProThr 343

 994 AGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCAT 1043
     ::::::||||||||:::::::::          :::|||   :::::
 344 LysArgGluLeuPheSerPheLeuArgIleGlyPheAsnLysProThrPh 360

1044 CACGCGTACGACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTTCA 1093
     :|||:::|||   |||:::   |||:::|||:::|||         :::|
 360 eThrLysThrTyrLeuSerGlyPhePheLysLysLysArg...ThrTyrT 376

1094 CG.........ACAGCCGTCAACGGTGGCGACCGCGCCATGGTGCCGATT 1134
       ||         |||   ::::::|||      |||   ::::::
 376 hrAsnProAspThrAsnLeuHisGlyGluThrArgProIleIleAspThr 392

1135 GGTACTTACGAGCGCGTAATGCCGCTAGACATCCTGCCTACCCTGCTTTT 1184
           |||::::::|||||||||:::   |||      :::   |||::
 393 AspIleTyrAspLysValMetProMetArgIleProValValProLeuIl 409

1185 GCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAAGCATTGGGTTGCT 1234
     ::::   :::|||:::   :::   |||   |||:::   ||||||   |
 409 eLysAlaValIleThrLysAsnPheAspLeuAlaAsnGluLeuGlyPheL 426

1235 TGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGC 1284
     |||||:::      |||||||:::||||||   :::::::::   |||:::
 426 euGluValCysGlyGluAspPheAlaLeuProThrLeuIleAspProSer 442

1285 AAATACGAATANGGCCCGCTGTTGCGTAAGGTGCTGGAAACCNTTGAGAA 1334
     |||   |||         ::::::::::::   |||            ||
 443 LysThrGluMetLeuThrIleValLysGluSerLeuIleGluTyrAlaLy 459

1335 GGAA 1338
     ||||
 459 sGlu 460
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY37553

seq_documentation_block:

```
ID   AAY37553 standard; Protein; 461 AA.
XX
AC   AAY37553;
XX
DT   07-OCT-1999  (first entry)
XX
DE   Chlamydia trachomatis surface exposed protein.
XX
KW   Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW   paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW   nongonococcal uretritis; epidymitis; cervicitis; salpingitis;
KW   bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX
OS   Chlamydia trachomatis.
XX
PN   WO9928475-A2.
XX
PD   10-JUN-1999.
XX
PF   27-NOV-1998;   98WO-IB01939.
XX
PR   04-NOV-1998;   98US-0107077.
PR   28-NOV-1997;   97FR-0015041.
PR   17-DEC-1997;   97FR-0016034.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-371125/31.
XX
PT   Genome sequence of Chlamydia trachomatis
XX
PS   Disclosure; Page 1216; 1755pp; English.
XX
CC   AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
CC   of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as
CC   vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC   can also be used to control growth of the microorganism. Chlamydia
CC   trachomatis is responsible for a large number of diseases, e.g. eye
CC   diseases such as conventional trachoma, nonendemic trachoma,
CC   paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC   nongonococcal uretritis, epidymitis, cervicitis, salpingitis,
CC   perihepatitis, bartholinitis; pneumopathy in breast feeding infants;
CC   and venereal lymphogranulomatosis. The polypeptides of the invention
CC   may be of use in treating these diseases.
XX
SQ   Sequence   461 AA;
```

alignment_scores:

```
            Quality:  432.00              Length:    461
              Ratio:  1.510                 Gaps:     13
 Percent Similarity:  62.039   Percent Identity: 28.200
```

alignment_block:

```
US-09-303-518D-127 x AAY37553   ..

Align seg 1/1  to: AAY37553  from: 1  to: 461

   4 ATTAAAATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCA 53
     |||   :::::::::||||||:::|||   :::   |||   |||:::::
   3 IleValValSerArgGlyLeuAspLeuSerLeuLysGlyAlaProLysGl 19

  54 AGTCATTTATGACGGGCCCGTC...ATTACCGAAGTCGCGTTGCTTGGCG 100
     :      :::   |||   |||      |||   |||::::::      :
  19 uSerGlyPheCysGlyLysValAspProThrTyrValSerValAspLeuA 36

 101 AAGAATATGCCGGTATGCGCCCCTNGATGAAAGTCAAGGAAGGCGATGCC 150
     ::   :::|||   :::         :::|||||      ||||||
  36 rgProPheAlaProLeuProLeuGluValLysValThrProGlyAspGln 52

 151 GTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGNATCCGGGCGTGGT 200
     |||      |||:::   |||   |||   |||          ||||||
  53 ValThrAlaGlySerProLeuAlaGluTyrLysLeuPheSerGlyValPh 69

 201 GTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCATCCATCGCGGCGAAA 250
      :::|||:::||||||:::|||:::::::::   |||:::||||||:::|
  69 eIleThrSerProValAspGlyGluValValGluIleArgArgGlyAsnL 86

 251 AGCGCGTACTTCAGTCGGTCGTGATTGCCGTTGAA...GGCAACGACGAA 297
     |||||:::|||   ::::::||||||      :::   |||   ::::::
  86 ysArgAlaLeuLeuGluIleValIleLysLysLysProGlyIleSerGln 102

 298 ATCGAGTTCGAACGCTACGCGCCCGAAGCGTTGGCAAACTTAAGCGGCGA 347
       :::|||:::                  |||   :::|||:::
 103 ThrLysPheSer.............TyrAspLeuGlnSerLeuThrGlnLy 115

 348 NGAANTNNGNNGCAATCTGATCCAATCCGGTTTGTGGACTGCGCTGCGTA 397
      :::            :::   ::::::||||||::::::   ::::::
 115 sAspLeuLeuGluValPheLysLysGluGlyLeuPheAlaLeuPheLysG 132

 398 NCCGTCCGTTCAGCAAAATCCCTGCCGTCGATGCCGAG...CCGTTCGCC 444
       |||||||||:::   |||||||||:::   ::::::   |||
 132 lnArgProPheAsp...IleProAlaLeuProThrGlnSerProArgAsp 147

 445 ATCTTCGTCAATGCGATGGACACCAATCCGCTN......GCGGCAGACCC 488
     :::|||:::|||      |||::::::|||:::      :::::::::
 148 ValPheIleAsnLeuAlaAspAsnArgProPheThrProSerValGluLy 164

 489 TGTGGTTGTGATCAAAGAAGCCGNCGANGAT.............TTCAGAC 526
       :::   :::   ::::::      |||             |||
 164 sHisLeuSerLeuPheSerSerLysGluAspGlyTyrTyrIlePheValV 181

 527 GANGTNTGCTGGTATTGAGCCGTTTGACCGAGCGTAAAATCCATGTGTGT 576
```

```
                 :::::::::::::|||          |||   |||:::
    181 alGlyValGlnAlaIleAlaLysLeuPheGlyLeuLysProHisIleIle 197

    577 AAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGCTGCCAACATC...GA 623
        ::::::            :::|||:::::::::   ::::::|||   ::
    198 SerThrAspArgLeuThrLeuProThrGlnAspLeuValSerIleAlaHi 214

    624 AACACATGAATTCGGCGGCCCGCATCCGGCCGGTTTGAGTGGCACGCACA 673
        :   |||   :::   ||||||   |||:::|||      :::|||||||
    214 sLeuHisThrIleAspGlyProPheProSerGlySerProSerThrHisI 231

    674 TTCATTTCATTGAGCCGGTC...GGTGCAAACAAAACCGTTTGGACCATC 720
        |||||   |||      :::   :::   :::   :::|||:::||||||
    231 leHisHisIleAlaArgIleArgAsnGluArgAspValValPheThrIle 247

    721 AATTATCAAGATGTAATTGCCATCGGACGTTTGTTTGCAACAGGCCGTCT 770
        :::::::|||:::|||:::::::||||||:::||||||      |||   ::
    248 SerPheGlnGluValLeuSerIleGlyHisLeuPheLeuLysGlyPheVa 264

    771 GAACACCGAGCGCGTGATTGCTTTGGGTGGTTCTCAAGTC......AACA 814
        :      ::::::::::::|||||||:::||||||   :::      ::::
    264 lLeuGlyGlnGlnIleValAlaLeuAlaGlySerAlaLeuProProSerG 281

    815 AACCACGCCTCTTGCGTACCGTTTTGGGTGCGAAAGTATCGCAAATTACT 864
        ::   :::   |||   |||:::   ||||||:::   |||::::::
    281 lnArgLysTyrLeuIleThrAlaLysGlyAlaSerPheSerAspLeuLeu 297

    865 GCGGGCGAATTGGTTGACGCAGACAAC...CGCGTGATTTCCGGTTCGGT 911
              ::::::   :::::::|||:::      :::||||||||:::
    298 ProLysAspIlePheSerSerAspGluIleThrLeuIleSerGlyAspPr 314

    912 ATTGAACGGCGCGATTACACAAGGCGCGCACGAT...TATTTGGGACGCT 958
         |||:::|||   :::   :::        ::::::      ||||||
    314 oLeuThrGlyArgLeuCysLysLysGluGluAsnProCysLeuGlyMetA 331

    959 ACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTC 1008
              :::   |||:::::::::   :::   ::::::::::|||   |||
    331 rgAspHisThrIleThrLeuLeuProAsnProLysThrArgGluSerPhe 347

   1009 GGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGTACGACCCT 1058
        :::::::::            :::|||   ::::::|||||||||   ||
    348 SerPheLeuArgLeuGlyTrpAsnLysLeuThrValThrArgThrTyrLe 364

   1059 CGGCCATTTCCTGAAAAACAAA......CTCTTCAAGTTCACGACAGCCG 1102
        |:::   |||:::|||:::|||      ::::::   :::   |||   :
    364 uSerGlyPhePheLysArgLysArgValPheMetAspMetAspThrAsnM 381

   1103 TCAACGGTGGCGACCGCGCCATGGTGCCGATTGGTACTTACGAGCGCGTA 1152
        :::::|||      |||   ::::::            ||||||||||||
    381 etHisGlyGluLysArgProIleIleAspAlaGluIleTyrGluArgVal 397

   1153 ATGCCGCTAGACATCCTGCCTACCCTGCTTTTGCGCGATTTAATCGTCGG 1202
             :::   :::      :::|||:::::::::   |||   :::
    398 SerAlaIleProValProValAlaLeuIleIleLysAlaLeuGluThrGl 414

   1203 CGATACCGACAGCGCGCAAGCATTGGGTTGCTTGGAATTGGACGAAGAAG 1252
         :::   :::::::|||      ||||||   |||||||::      ||||
    414 nAsnPheGluGluAlaCysArgLeuGlyLeuLeuGluValAlaProGluA 431

   1253 ACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATANGGCCCG 1302
        ||:::||||||   :::|||:::   |||:::|||   |||
    431 spPheAlaLeuProThrPheIleAspProSerLysThrGluMetPheSer 447

   1303 CTGTTGCGTAAGGTGCTGGAAACCNTTGAGAAG 1335
        ::::::::::::   |||      ::::::|||
    448 IleValLysGluSerLeuLeuArgThrGlnLys 458

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34467

seq_documentation_block:
ID   AAY34467 standard; Protein; 443 AA.
XX
AC   AAY34467;
XX
DT   25-AUG-1999  (first entry)
XX
DE   Porphorymonas gingivalis protein PG122.
XX
KW   Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
KW   vaccine; antigenic.
XX
OS   Porphorymonas gingivalis.
XX
PN   WO9929870-A1.
XX
PD   17-JUN-1999.
XX
PF   10-DEC-1998;   98WO-AU01023.
XX
PR   04-AUG-1998;   98AU-0005028.
PR   10-DEC-1997;   97AU-0000839.
PR   31-DEC-1997;   97AU-0001182.
PR   30-JAN-1998;   98AU-0001546.
PR   10-MAR-1998;   98AU-0002264.
PR   09-APR-1998;   98AU-0002911.
PR   23-APR-1998;   98AU-0003128.
PR   05-MAY-1998;   98AU-0003338.
PR   22-MAY-1998;   98AU-0003654.
PR   29-JUL-1998;   98AU-0004917.
XX
PA   (CSLC-) CSL LTD.
XX
PI   Agius CT,  Barr IG,  Hocking DM,  Margetts MB,  Patterson MA;
PI   Ross BC,  Rothel LJ,  Webb EA;
XX
DR   WPI; 1999-385613/32.
DR   N-PSDB; AAX91685.
XX
PT   Antigenic Porphorymonas gingivalis peptides for preventing
PT   gingivitis
XX
PS   Claim 1; Page 445-446; 588pp; English.
XX
CC   AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC   Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC   AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC   isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC   activity with a vaccine mechanism of action. The PG polypeptides can be
CC   used as vaccines especially against Porphorymonas gingivalis. Probes can
CC   be used to detect Porphorymonas gingivalis in standard hybridisation
CC   assays. Porphorymonas gingivalis is involved in periodontal disease
CC   especially gingivitis.
XX
SQ   Sequence   443 AA;


alignment_scores:
            Quality:  159.50              Length:    469
              Ratio:  0.697                 Gaps:     17
 Percent Similarity:  48.827   Percent Identity: 19.616

alignment_block:
 US-09-303-518D-127 x AAY34467   ..

 Align seg 1/1  to: AAY34467  from: 1  to: 443

      37 GCGGGCAGACCGGAGCAAGTCATTTATGACGGGCCCGTCATTACCGAAGT 86
         ||||||:::|||   :::|||:::          |||:::   :::::::||
      19 AlaGlyLysProValGluValLeu.........ProIleProSerGlnVa 32

      87 CGCGTTGCTTGGCGAAGAATATGCCGGTATGCGCCCCTNGATGAAAGTCA 136
         |:::::          ::::::   |||          :::      ||||
      32 lValIleProLeuGlyGlnHisIleGlyAlaProAlaThrAlaThrValL 49

     137 AGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAG 186
         ||:::|||||   ||||||   |||   ::::::   :::
```

```
 49 ysLysGlyAspGluValLysValGlyThrIleIleAlaGlnAlaGlyGly 65

187 NATCCGGGCGTGGTGTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCAT 236
        ::::::   :::   :::   ||||||||||||::      ||
 66 PheValSerAlaAsnIleHisSerSerValSerGlyLysValLeuLysIl 82

237 C.........CATCGCGGCGAAAAGCGCGTACTTCAGTCGGTCGTGATTG 277
    |         :::   ::::::                ::::||   |||:
 82 eAspAsnValTyrAspSerSerGlyTyrProLysProAlaValPheIleS 99

278 CCGTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGCGCCCGAAGCG 327
    ::||||||||||::::::      |||             :::|||
 99 erValGluGlyAspGluTrpGluGluGlyIleAspArgSerProAlaIle 115

328 TTG......GCAAACTTAAGCGGCGANGAANTNNGNNGCAATCTGATCCA 371
    :::      :::||||||::::::   |||      :::::::::
116 ValLysGluCysAsnLeuAspAlaLysGluIleValAlaLysIleSerAl 132

372 ATCCGGTTTGTGGACTGCGCTGCGTANCCGTCCGTTC..............  408
     :::|||:::   :::::::|||           |||
132 aAlaGlyIle...ValGlyLeuGlyGlyAlaThrPheProThrHisValL 148

409 .....AGCAAAATCCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTCGTC 453
        |||      |||:::      ||||||         :::::::::
148 ysLeuSerProProProGlyAsnLysAlaGlu......IleLeuIleIle 162

454 AATGCGATGGACACCAATCCG...CTNGCGGCAGACCCTGTGGTTGTGAT 500
    ||||||::::::   :::|||   |||::::::|||   |||:::::::::
163 AsnAlaValGluCysGluProTyrLeuThrSerAspHisValLeuMetLe 179

501 CAAAGAAGCCGNCGANGATTTCAGACGANGTNTGCTGGTATTGAGCCGTT 550
    ::::::::::          :::         ::::::::::::::    :
179 uGluHisGlyGluGluIleMetIleGlyValSerIleLeuMetLysAlaI 196

551 TGACCGAGCGTAAA..................................... 564
    ::      :::|||
196 leGlnValAsnLysAlaValIleGlyValGluAsnAsnLysLysAspAla 212

565 ...ATCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGC 611
       |||:::   |||   |||   |||
213 IleAlaHisLeuThrLysLeuAlaThrAla.................Ty 223

612 TGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCGGCCGGTTTGA 661
       :::|||||||:::      :::         :::|||   |||   :
223 rProGlyIleGluValMetProLeuLysValGlnTyrProGlnGlyGlyG 240

662 GTGGCACGCACATTCATTTCATT............................ 684
    ::         |||      :::
240 luLysGlnLeuIleAspAlaValIleArgLysGlnValLysSerGlyAla 256

685 GAGCCGGTCGGTGCAAACAAAACCGTTTGGACCATCAATTATCAAGATGT 734
       |||:::::::::::::   :::|||         |||         ||
257 LeuProIleSerThrGlyAlaValValGln......AsnValGlyThrVa 271

735 AATTGCCATCGGACGTTTGTTTGCAACAGGCCGTCTGAACACCGAGCGCG 784
    |:::|||:::   :::          ::::::      :::|||||||:
271 lPheAlaValTyrGluAlaValGlnLysAsnLysProLeuValGluArgI 288

785 TGATTGCTTTGGGTGGTTCTCAAGTCAACAAACCACGCCTCTTGCGTACC 834
    ::::::::::   |||:::::::::::::::|||      |||   :::
288 leValThrValThrGlyLysLysLeuSerArgProSerAsnLeuLeuVal 304

835 GTTTTGGGTGCGAAAGTATCGCAAATT......ACTGCGGGCGAATTGGT 878
       :::|||:::   ::::::   :::      :::||||||   |||
305 ArgIleGlyThrProIleAlaAlaLeuIleGluAlaAlaGlyGlyLeuPr 321

879 TGACGCAGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGCGCGATTA 928
     :::      :::::::::|||:::|||:::
321 oGluAsnThrGlyLysIleIleGlyGlyGly.................. 331

929 CACAAGGCGCGCACGATTATTTGGGACGCTACCACAATCAGATTTCCGTT 978

331 ..................................................  331

979 ATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCC 1028
                                  ::::::|||   :::      :::|||
332 .....................ProMetMetGlyArgAlaLeuLeuSerPr 341

1029 GGACAAATACTCCATCACGCGTACGACCCTCGGCCATTTCCTGAAAAACA 1078
     ||||         :::|||:::   :::   |||   ::::::   ::::
 341 oAsp...ValProValThrLysGlySerSerGlyValLeuIleLeuAspA 357

1079 AACTCTTCAAGTTCACGACAGCCGTCAACGGTGGCGACCGCGCCATGGTG 1128
     ::                  ||||||:::          |||      :::
 357 rg.............GluGluAlaValArgLysProMetArgAspCysIle 369

1129 CCGATTGGTACTTACGAGCGCGTAATGCCGCTAGACATCCTGCCTACCCT 1178
     :::                 |||   |||:::   :::   |||:::::
 370 ArgCysAlaLysCysValGlyValCysProMetGlyLeuAsnProAlaPh 386

1179 GCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAA....... 1221
     :|||:::||||||   :::      :::   ::::::|||:::
 386 eLeuMetArgAspThrLeuTyrLysSerTrpGluThrAlaGluLysGlyA 403

1222 ..GCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGC 1269
       ::::::   |||:::|||      :::                |||||
 403 snValValAspCysIleGluCysGlySer.............CysSer 414

1270 TTCGTCTGCCCGGGCAAATACGAATANGGCCCGCTGTTGCGTAAGGTGCT 1319
     |||:::||||||::::::                  :::|||::::::
 415 PheThrCysProAlaAsnArgProLeuLeuAspTyrIleArgGlnAlaLy 431

1320 GGAAACC 1326
      :::|||
 431 sLysThr 433
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY34343

seq_documentation_block:

```
ID   AAY34343 standard; Protein; 451 AA.
XX
AC   AAY34343;
XX
DT   25-AUG-1999  (first entry)
XX
DE   Porphorymonas gingivalis protein PG122.
XX
KW   Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
KW   vaccine; antigenic.
XX
OS   Porphorymonas gingivalis.
XX
PN   WO9929870-A1.
XX
PD   17-JUN-1999.
XX
PF   10-DEC-1998;   98WO-AU01023.
XX
PR   04-AUG-1998;   98AU-0005028.
PR   10-DEC-1997;   97AU-0000839.
PR   31-DEC-1997;   97AU-0001182.
PR   30-JAN-1998;   98AU-0001546.
PR   10-MAR-1998;   98AU-0002264.
PR   09-APR-1998;   98AU-0002911.
PR   23-APR-1998;   98AU-0003128.
PR   05-MAY-1998;   98AU-0003338.
PR   22-MAY-1998;   98AU-0003654.
PR   29-JUL-1998;   98AU-0004917.
XX
PA   (CSLC-) CSL LTD.
XX
PI   Agius CT,  Barr IG,  Hocking DM,  Margetts MB,  Patterson MA;
PI   Ross BC,  Rothel LJ,  Webb EA;
XX
DR   WPI; 1999-385613/32.
DR   N-PSDB; AAX91561.
```

```
XX
PT   Antigenic Porphorymonas gingivalis peptides for preventing
PT   gingivitis
XX
PS   Claim 1; Page 303; 588pp; English.
XX
CC   AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC   Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC   AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC   isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC   activity with a vaccine mechanism of action. The PG polypeptides can be
CC   used as vaccines especially against Porphorymonas gingivalis. Probes can
CC   be used to detect Porphorymonas gingivalis in standard hybridisation
CC   assays. Porphorymonas gingivalis is involved in periodontal disease
CC   especially gingivitis.
XX
SQ   Sequence   451 AA;


alignment_scores:
           Quality:  159.50            Length:    469
             Ratio:  0.697              Gaps:     17
 Percent Similarity:  48.827   Percent Identity: 19.616

alignment_block:
 US-09-303-518D-127 x AAY34343   ..

 Align seg 1/1  to: AAY34343  from: 1  to: 451

      37 GCGGGCAGACCGGAGCAAGTCATTTATGACGGGCCCGTCATTACCGAAGT 86
         ||||||:::|||   :::|||:::         |||:::   :::::::||
      27 AlaGlyLysProValGluValLeu.........ProIleProSerGlnVa 40

      87 CGCGTTGCTTGGCGAAGAATATGCCGGTATGCGCCCCTNGATGAAAGTCA 136
         |:::::         ::::::   |||         :::      ||||
      40 lValIleProLeuGlyGlnHisIleGlyAlaProAlaThrAlaThrValL 57

     137 AGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAG 186
         ||:::||||||   ||||||   |||   ::::::   :::
      57 ysLysGlyAspGluValLysValGlyThrIleIleAlaGlnAlaGlyGly 73

     187 NATCCGGGCGTGGTGTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCAT 236
               ::::::   :::   :::   ||||||||||||:::      ||
      74 PheValSerAlaAsnIleHisSerSerValSerGlyLysValLeuLysIl 90

     237 C.........CATCGCGGCGAAAAGCGCGTACTTCAGTCGGTCGTGATTG 277
         |         :::   ::::::                ::::|||   |||:
      90 eAspAsnValTyrAspSerSerGlyTyrProLysProAlaValPheIleS 107

     278 CCGTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGCGCCCGAAGCG 327
         ::|||||||||::::::      |||             :::|||
     107 erValGluGlyAspGluTrpGluGluGlyIleAspArgSerProAlaIle 123

     328 TTG......GCAAACTTAAGCGGCGANGAANTNNGNNGCAATCTGATCCA 371
         :::      :::|||||||::::::   |||      :::::::::
     124 ValLysGluCysAsnLeuAspAlaLysGluIleValAlaLysIleSerAl 140

     372 ATCCGGTTTGTGGACTGCGCTGCGTANCCGTCCGTTC.............408
          :::|||:::   :::::::|||            |||
     140 aAlaGlyIle...ValGlyLeuGlyGlyAlaThrPheProThrHisValL 156

     409 .....AGCAAAATCCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTCGTC 453
              |||      |||:::      ||||||         :::::::::
     156 ysLeuSerProProProGlyAsnLysAlaGlu......IleLeuIleIle 170

     454 AATGCGATGGACACCAATCCG...CTNGCGGCAGACCCTGTGGTTGTGAT 500
         ||||||::::::   :::|||   |||:::::|||   |||:::::::::
     171 AsnAlaValGluCysGluProTyrLeuThrSerAspHisValLeuMetLe 187

     501 CAAAGAAGCCGNCGANGATTTCAGACGANGTNTGCTGGTATTGAGCCGTT 550
         ::::::::::         :::          :::::::::::::::   :
     187 uGluHisGlyGluGluIleMetIleGlyValSerIleLeuMetLysAlaI 204

     551 TGACCGAGCGTAAA..................................... 564
         ::      :::|||
     204 leGlnValAsnLysAlaValIleGlyValGluAsnAsnLysLysAspAla 220

     565 ...ATCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGC 611
            |||:::   |||   |||   |||
     221 IleAlaHisLeuThrLysLeuAlaThrAla.................Ty 231

     612 TGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCGGCCGGTTTGA 661
         :::|||||||:::      :::         :::|||   |||   :
     231 rProGlyIleGluValMetProLeuLysValGlnTyrProGlnGlyGlyG 248

     662 GTGGCACGCACATTCATTTCATT........................... 684
         ::         |||      :::
     248 luLysGlnLeuIleAspAlaValIleArgLysGlnValLysSerGlyAla 264

     685 GAGCCGGTCGGTGCAAACAAAACCGTTTGGACCATCAATTATCAAGATGT 734
         |||:::::::::::::   :::|||         |||         ||
     265 LeuProIleSerThrGlyAlaValValGln......AsnValGlyThrVa 279

     735 AATTGCCATCGGACGTTTGTTTGCAACAGGCCGTCTGAACACCGAGCGCG 784
         |:::|||:::   :::                 ::::::      :::||||||:
     279 lPheAlaValTyrGluAlaValGlnLysAsnLysProLeuValGluArgI 296

     785 TGATTGCTTTGGGTGGTTCTCAAGTCAACAAACCACGCCTCTTGCGTACC 834
         ::::::::::   |||:::::::::::::::|||      |||   :::
     296 leValThrValThrGlyLysLysLeuSerArgProSerAsnLeuLeuVal 312

     835 GTTTTGGGTGCGAAAGTATCGCAAATT......ACTGCGGGCGAATTGGT 878
            :::|||:::   ::::::   :::      :::||||||   |||
     313 ArgIleGlyThrProIleAlaAlaLeuIleGluAlaAlaGlyGlyLeuPr 329

     879 TGACGCAGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGCGCGATTA 928
         :::      :::::::::|||:::|||:::
     329 oGluAsnThrGlyLysIleIleGlyGlyGly................... 339

     929 CACAAGGCGCGCACGATTATTTGGGACGCTACCACAATCAGATTTCCGTT 978

     339 ..................................................... 339

     979 ATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCC 1028
                                    ::::::|||   :::      :::||
     340 .....................ProMetMetGlyArgAlaLeuLeuSerPr 349

    1029 GGACAAATACTCCATCACGCGTACGACCCTCGGCCATTTCCTGAAAAACA 1078
         ||||         :::|||:::   :::   |||   ::::::   ::::
     349 oAsp...ValProValThrLysGlySerSerGlyValLeuIleLeuAspA 365

    1079 AACTCTTCAAGTTCACGACAGCCGTCAACGGTGGCGACCGCGCCATGGTG 1128
         ::                  |||||||:::         |||      :::
     365 rg............GluGluAlaValArgLysProMetArgAspCysIle 377

    1129 CCGATTGGTACTTACGAGCGCGTAATGCCGCTAGACATCCTGCCTACCCT 1178
                :::            |||   |||:::   :::   |||:::::
     378 ArgCysAlaLysCysValGlyValCysProMetGlyLeuAsnProAlaPh 394

    1179 GCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAA....... 1221
         :|||:::||||||   :::      :::   ::::::|||:::
     394 eLeuMetArgAspThrLeuTyrLysSerTrpGluThrAlaGluLysGlyA 411

    1222 ..GCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGC 1269
           ::::::   |||:::|||      :::                ||||||
     411 snValValAspCysIleGluCysGlySer..............CysSer 422

    1270 TTCGTCTGCCCGGGCAAATACGAATANGGCCCGCTGTTGCGTAAGGTGCT 1319
         |||:::||||||::::::                     :::|||::::::
     423 PheThrCysProAlaAsnArgProLeuLeuAspTyrIleArgGlnAlaLy 439

    1320 GGAAACC 1326
          :::|||
     439 sLysThr 441

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59813
```

```
seq_documentation_block:
ID   AAB59813 standard; Protein; 1017 AA.
XX
AC   AAB59813;
XX
DT   04-APR-2001  (first entry)
XX
DE   TutD protein #4.
XX
KW   Toluene degradation; enzyme; waste degradation; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23625, AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs  -
XX
PS   Disclosure; Fig 5; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence for toluene degrading enzyme, TutD.
XX
SQ   Sequence   1017 AA;


alignment_scores:
            Quality:  134.50              Length:    469
              Ratio:  0.659                Gaps:     22
 Percent Similarity:  43.497   Percent Identity: 22.175

alignment_block:
 US-09-303-518D-127 x AAB59813   ..

 Align seg 1/1  to: AAB59813  from: 1  to: 1017

      35 TCGCGGGCAGACCGGAGCAAGTCATTTATGACGGGCCCGTCATTACCGAA 84
         |||||||||      :::||||||                      ::
     260 SerArgAlaProThrAlaLysSer.....................SerAr 269

      85 GTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGCCCCTNGATGAAAGT 134
         ::::|||   ::::::::::::   |||   :::|||||||:::   ::::
     269 gGlyArgThrIleCysSerSerSerProSerAlaAlaProThrProArgA 286

     135 CAAGG..................................................... 139
         ::|||
     286 laArgThrProAlaThrThrProThrProSerSerArgGlnProSerGly 302

     140 AAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGNAT 189
         :::|||   ||||||      ::::::   ::::::      ::::::
     303 SerAlaArgProSerProProSerSerSerAlaIleProArgArgThrAl 319

     190 CCGGGCGTGGTGTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCATCCA 239
          |||      |||           |||::::::   |||      :::
     319 aArgArgArgCys......AlaGlyPheSerSerAlaSerAlaThrAspS 334

     240 TCGCGGCGAAAAGCGCGTACTTCAGTCGGTCGTGATTGCCGTTGAAGGCA 289
         |||   :::   :::      :::||||||                :::
     334 erAlaIleArgArgSerSerThrThrArgSerAlaArgSerArgArgAsn 350

     290 ACGACGAAATCGAGTTCGAACGCTACGCGCCCGAAGCGTTGGCAAACTTA 339
         |||   :::|||:::|||:::||||||   |||
     351 ThrProSerSerAlaSerThrAlaThrAlaProProThr.......... 363

     340 AGCGGCGANGAANTNNGNNGCAATCTGATCCAATCCGGTTTGTGGACTGC 389
                                        :::|||:::   |||
     364 ............................ArgLysProThrThrGlySerT 371

     390 GCTGCGTANCCGTCCGTTCAGCAAAATCCCTGCCGTCGATGCCGAGCCGT 439
         |||   :::   |||   |||::::::::
     371 hrCysCysAlaCysArgProAlaSerThrVal.................. 381

     440 TCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTNGCGGCAGACCCT 489
                :::::   |||      |||:::|||
     382 .........AlaAlaArgArgLysLysProValArg.............. 390

     490 GTGGTTGTGATCAAAGAAGCCGNCGANGATTTCAGACGANGTNTGCTGGT 539
                      |||      :::      |||      ||||||
     391 .............LysValAlaAlaGlnSerSerArgProSerCysTrp. 402

     540 ATTGAGCCGTTTGACCGAGCGTAAAATCCATGTGTGTAAGGCAGCTGGCG 589
                                                :::::   :::
     403 ....................................LysSerArgSer 406

     590 CAGACGTGCCGTCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGC 639
         :::|||:::                |||||                ||||||||
     407 MetThrAlaThrThrGlyArgThrProThr.........CysAsnSerAl 420

     640 GGCCCGCATCCGGCCGGTTTGAGTGGCACGCACATTCATTTCATTGAGCC 689
         |         |||||||||         |||:::         :::
     420 aArg......ArgProValIleSerArgArgSerProSerArgMet.Phe 434

     690 GGTCGGTGCAAACAAAACCGTTTGGACCATCAATTATCAAGATGTAATTG 739
         ||||||   :::   :::   ::::::   :::   :::      :::::
     435 GlyArgLeuSerAlaSerSerIleAsnMetArgSerThrSerValSerAl 451

     740 CCATCGGACGTTTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTGATT 789
         :   ||||||      |||:::      |||:::      :::      |
     451 aProArgThr......CysArgAlaThrSerSerSerAlaSerCysArgC 466

     790 GCTTTGGGTGGTTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTT 839
         ||         :::|||      |||:::|||      |||::::::
     466 ys.........LeuSerCysProGlnSerThrThrAlaAlaTrpAsnSer 479

     840 GGG..............TGCGAAAGTATCGCAAATTACTGCGGGCGAAT 874
         |||               |||   |||   :::      :::|||
     480 GlyTrpThrProAlaProCysProSer...SerProMetAlaGlyThrTh 495

     875 TGGTTGACGCAGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGCGCG 924
                |||||:::   |||                  :::|||:::
     495 rArgSerArgArgSerSerArgArgThrProSerTrpProSerArgAsnT 512

     925 ATTACACAAGGCGCGCACGATTATTTGGGACGCTACCACAATCAGATTTC 974
         |||:::||||||              :::   |||:::|||:::
     512 rpTyrSerArgArgArg............AsnThrProSerSerAsnSer 524

     975 CGTTATCGAAGAAGGCC.......GCAGCAAAGAGCTGT.........TC 1008
               |||||||||          :::::::::::|||          ||
     525 AlaLysArgArgArgThrGlyLysValSerArgLysCysAlaSerThrSe 541

    1009 GGCTGGGTTGCGCCGCAGCCGGACAAAT......ACTCCATCACGCGTAC 1052
         |:::|||   |||   :::   |||:::      ||||||:::
```

```
     541 rSerGlyArgArgSerGlyAlaThrThrMetIleThrProThrValSerS 558

    1053 GACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTTCACGACAGCCG 1102
         |||::::::         ::::::   ::::::   :::|||:::|||
     558 erProAlaSerThrArgLysSerSerAlaAlaLysCysAlaArgSerPro 574

    1103 TCAACGGTGGCGACCGCGCCATGGTGCCGATTGGTACTTACGAGCGCGTA 1152
         :::|||:::::::::         ||||||||||||      ||||||
     575 ThrThrLeuValValArgSer...CysArgLeuValArgLeuSerAla.. 589

    1153 ATGCCGCTAGACATCCTGCCTACCCTGCTTTTGCGCGATTTAATCGTCGG 1202
                                      |||:::      |||
     590 ..............................CysThrTrpLysSerValA 596

    1203 CGATACCGACAGCGCGCAAGCATTGGGTTGCTTGGAATTGGACGAAGAAG 1252
                :::   |||:::
     596 rgAlaArgAlaProArgArg............................. 602

    1253 ACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATANGGCC.. 1300
                      :::||||||:::::::::   :::|||::::::|||
     603 .........ThrGlyAlaSerGlyValLysArgGlnThrThrAlaAlaPh 616

    1301 ....................................CGCTGTTGCGTAA 1313
                                             |||||||||:::|
     616 eLeuProThrTrpGluProThrArgArgGlyArgArgArgCysCysAlaA 633

    1314 GGTGC 1318
         |||||
     633 rgCys 634
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826

seq_documentation_block:

ID AAB59826 standard; Protein; 1615 AA.
XX
AC AAB59826;
XX
DT 04-APR-2001 (first entry)
XX
DE Protein #3 encoded by TutD/E gene.
XX
KW Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS Thauera aromatica.
OS Xanthomonas maltophilia.
OS Geobacter metallireducens.
OS Azoarcus tolulyticus.
XX
PN WO200072650-A2.
XX
PD 07-DEC-2000.
XX
PF 24-MAY-2000; 2000WO-US14298.
XX
PR 01-JUN-1999; 99US-0323872.
XX
PA (UYOH-) UNIV OHIO.
XX
PI Coschigano PW;
XX
DR WPI; 2001-041080/05.
DR N-PSDB; AAF23627.
XX
PT Composition comprising toluene degrading enzyme useful for biological
PT treatment of organic compounds, especially for degrading toluene or its
PT analogs -
XX
PS Disclosure; Fig 12; 122pp; English.
XX
CC The present invention relates to toluene degrading enzyme genes and
CC proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC toluene degrading enzymes are useful for biological treatment of organic
CC compounds and in particular for the degradation of toluene and its
CC analogs contained in liquid or solid waste source. The present sequence
CC is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
XX
SQ Sequence 1615 AA;

alignment_scores:
Quality: 134.50 Length: 469
Ratio: 0.659 Gaps: 22
Percent Similarity: 43.497 Percent Identity: 22.175

alignment_block:
US-09-303-518D-127 x AAB59826 ..

Align seg 1/1 to: AAB59826 from: 1 to: 1615

```
      35 TCGCGGGCAGACCGGAGCAAGTCATTTATGACGGGCCCGTCATTACCGAA 84
         |||||||||      :::|||||||                      ::
     858 SerArgAlaProThrAlaLysSer.....................SerAr 867

      85 GTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGCCCCTNGATGAAAGT 134
         ::::|||   ::::::::::::   |||   :::|||||||:::   ::::
     867 gGlyArgThrIleCysSerSerSerProSerAlaAlaProThrProArgA 884

     135 CAAGG..................................................... 139
         ::|||
     884 laArgThrProAlaThrThrProThrProSerSerArgGlnProSerGly 900

     140 AAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGNAT 189
         :::|||   ||||||      ::::::   ::::::      ::::::
     901 SerAlaArgProSerProProSerSerSerAlaIleProArgArgThrAl 917

     190 CCGGGCGTGGTGTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCATCCA 239
          |||      |||            |||::::::   |||      :::
     917 aArgArgArgCys......AlaGlyPheSerSerAlaSerAlaThrAspS 932

     240 TCGCGGCGAAAAGCGCGTACTTCAGTCGGTCGTGATTGCCGTTGAAGGCA 289
         |||   :::   :::      :::||||||           :::
     932 erAlaIleArgArgSerSerThrThrArgSerAlaArgSerArgArgAsn 948

     290 ACGACGAAATCGAGTTCGAACGCTACGCGCCCGAAGCGTTGGCAAACTTA 339
         |||   :::|||:::|||:::||||||   |||
     949 ThrProSerSerAlaSerThrAlaThrAlaProProThr.......... 961

     340 AGCGGCGANGAANTNNGNNGCAATCTGATCCAATCCGGTTTGTGGACTGC 389
                                      :::|||:::   |||
     962 ..........................ArgLysProThrThrGlySerT 969

     390 GCTGCGTANCCGTCCGTTCAGCAAAATCCCTGCCGTCGATGCCGAGCCGT 439
         |||   :::   |||   |||:::::::::
     969 hrCysCysAlaCysArgProAlaSerThrVal.................. 979

     440 TCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTNGCGGCAGACCCT 489
                 ::::::   |||      |||:::|||
     980 .........AlaAlaArgArgLysLysProValArg.............. 988

     490 GTGGTTGTGATCAAAGAAGCCGNCGANGATTTCAGACGANGTNTGCTGGT 539
                    |||      :::      |||         ||||||
     989 .............LysValAlaAlaGlnSerSerArgProSerCysTrp. 1000

     540 ATTGAGCCGTTTGACCGAGCGTAAAATCCATGTGTGTAAGGCAGCTGGCG 589
                                                  ::::::   :::
    1001 .....................................LysSerArgSer 1004

     590 CAGACGTGCCGTCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGC 639
         :::|||:::                ||||||           ||||||||
    1005 MetThrAlaThrThrGlyArgThrProThr.........CysAsnSerAl 1018

     640 GGCCCGCATCCGGCCGGTTTGAGTGGCACGCACATTCATTTCATTGAGCC 689
         |         |||||||||         |||:::      :::
    1018 aArg......ArgProValIleSerArgArgSerProSerArgMet.Phe 1032
```

```
 690 GGTCGGTGCAAACAAAACCGTTTGGACCATCAATTATCAAGATGTAATTG 739
     ||||||   :::   :::   ::::::   :::   :::      :::::
1033 GlyArgLeuSerAlaSerSerIleAsnMetArgSerThrSerValSerAl 1049

 740 CCATCGGACGTTTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTGATT 789
     :  ||||||      |||:::      |||:::      :::       |
1049 aProArgThr......CysArgAlaThrSerSerSerAlaSerCysArgC 1064

 790 GCTTTGGGTGGTTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTT 839
     ||         :::|||      |||:::|||      |||::::::
1064 ys.........LeuSerCysProGlnSerThrThrAlaAlaTrpAsnSer 1077

 840 GGG..............TGCGAAAGTATCGCAAATTACTGCGGGCGAAT 874
     |||                |||   |||   :::      :::|||
1078 GlyTrpThrProAlaProCysProSer...SerProMetAlaGlyThrTh 1093

 875 TGGTTGACGCAGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGCGCG 924
           ||||||:::   |||                  :::|||:::
1093 rArgSerArgArgSerSerArgArgThrProSerTrpProSerArgAsnT 1110

 925 ATTACACAAGGCGCGCACGATTATTTGGGACGCTACCACAATCAGATTTC 974
       |||:::||||||                 :::   |||:::|||::
1110 rpTyrSerArgArgArg............AsnThrProSerSerAsnSer 1122

 975 CGTTATCGAAGAAGGCC.......GCAGCAAAGAGCTGT.........TC 1008
           |||||||||          :::::::::::::|||         ||
1123 AlaLysArgArgArgThrGlyLysValSerArgLysCysAlaSerThrSe 1139

1009 GGCTGGGTTGCGCCGCAGCCGGACAAAT......ACTCCATCACGCGTAC 1052
     |:::|||   |||   :::   |||:::      |||||||::
1139 rSerGlyArgArgSerGlyAlaThrThrMetIleThrProThrValSerS 1156

1053 GACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTTCACGACAGCCG 1102
       |||::::::         ::::::   ::::::   :::|||:::|||
1156 erProAlaSerThrArgLysSerSerAlaAlaLysCysAlaArgSerPro 1172

1103 TCAACGGTGGCGACCGCGCCATGGTGCCGATTGGTACTTACGAGCGCGTA 1152
     :::|||:::::::::         ||||||||||||      ||||||
1173 ThrThrLeuValValArgSer...CysArgLeuValArgLeuSerAla.. 1187

1153 ATGCCGCTAGACATCCTGCCTACCCTGCTTTTGCGCGATTTAATCGTCGG 1202
                                  |||:::      |||
1188 ..............................CysThrTrpLysSerValA 1194

1203 CGATACCGACAGCGCGCAAGCATTGGGTTGCTTGGAATTGGACGAAGAAG 1252
           :::   |||:::
1194 rgAlaArgAlaProArgArg.............................. 1200

1253 ACCTCGCTTTGTGCAGCTTCGTCTGCCCGGGCAAATACGAATANGGCC.. 1300
               :::|||||||:::::::::   :::|||:::::|||
1201 .........ThrGlyAlaSerGlyValLysArgGlnThrThrAlaAlaPh 1214

1301 ....................................CGCTGTTGCGTAA 1313
                                         |||||||||:::|
1214 eLeuProThrTrpGluProThrArgArgGlyArgArgArgCysCysAlaA 1231

1314 GGTGC 1318
     |||||
1231 rgCys 1232
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59817

seq_documentation_block:

```
ID   AAB59817 standard; Protein; 999 AA.
XX
AC   AAB59817;
XX
DT   04-APR-2001  (first entry)
XX
DE   TutD protein #8.
XX
KW   Toluene degradation; enzyme; waste degradation; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23625, AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs -
XX
PS   Disclosure; Fig 5; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence for toluene degrading enzyme, TutD.
XX
SQ   Sequence   999 AA;
```

alignment_scores:

```
           Quality:  132.00              Length:    504
             Ratio:  0.706                Gaps:     25
Percent Similarity:  37.103   Percent Identity: 22.619
```

alignment_block:

US-09-303-518D-127 x AAB59817 ..

Align seg 1/1 to: AAB59817 from: 1 to: 999

```
  81 CGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGCCCCTNGATGA 130
     |||||||||   ::       |||   ||||||::::::      :::::
 117 ArgSerArgCysSerProAspArgCysCysArgTrpSerArgCysSerSe 133

 131 AAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 180
     :   :::   |||||||||      :::   |||::::::
 133 rProSerProArgArgCysProProSerSerProAlaGly.......... 146

 181 AAAAAGNATCCGGGCGTGGTGTTTACCGCGCCNGTTTCAGGCAAAATCGC 230
           :::   |||           |||      |||   ::::::|||
 147 .....AlaProGlyAlaThrCysSerArgArgProPheSerArgSerArg 161

 231 CGCCATCCATCGCGGCGAAAAGCGCGTACTTCAGTCGGTCGTGAT...TG 277
               :::   |||   ||||||            ||||||   ||
 162 AspSerAlaGlyProArgAlaAlaArgPheArgArgCysArgAspAlaCy 178

 278 CCGTTGAAGGCAACGA................................ 293
     |:::   |||   |||
 178 sGluArgArgAlaArgCysProGlyProArgSerAlaProSerIleArgA 195

 294 ........CGAAATCGAGTTCGAACGCTACGCGCCCGAAGCGTTGGCAAA 335
             |||:::|||   |||:::   |||:::||||||   |||:::
 195 rgGlySerArgAspArgSerArgAlaSerArgSerArgSerArgGlySer 211
```

```
 336 ..............................CTTAAGCGGCGANGAAN 352
                                  :::|||||||   ::
 212 ProLeuCysGlyAlaThrAlaThrSerCysProArgArgArgArgCysSe 228

 353 TNNGNNGCAATCTGATCCAATCCGGTTTGTGGACTGCGCTGCGTANCCGT 402
     :      |||:::                :::      |||
 228 rIleGlyAlaSerSerGly.................CysProHisProP 239

 403 CCGTTCAGCAAAATCCCTGC.....................CGTCGATGC 431
       |||::::::::::|||                      |||||||||
 239 roValArgArgSerProValAsnSerSerLysArgAlaHisArgArgCys 255

 432 CGAGCCGTTCGC...............................CATCT 448
        |||   |||
 256 ThrAlaArgArgGlyArgPheArgGlyProThrSerArgAspThrGlyAr 272

 449 TCGTCAATGCGATGGACACCAATCCGCTNGCGGCAGACCCTGTGGTTGTG 498
      |||:::|||         :::            |||   |||   |||:
 272 gArgArgCys......TrpArgTrpProArgProArgArgCysArgCysS 287

 499 ATCAAAGAAGCCGNCGANGATTTCAGACGANGTNTGCTGGTATTGAGCCG 548
     :::::|||      |||   :::   :::
 287 erArgArgTrpGlyArgProLeuTrpAla..................... 296

 549 TTTGACCGAGCGTAAAATCCATGTGTGTAAGGCAGCTGGCGCAGA.....593
                    :::   |||      :::   |||||||||
 297 ..............SerGlyCysProArgAlaArgTrpArgArgGlySe 308

 594 ............................CGTGCCGTCTGAAAATGCTGCC 615
                                 :::      :::   |||
 308 rAsnTrpSerSerGlyArgSerSerAlaAlaSerProLysArgThrCysG 325

 616 AACATCGA.........AACACATGAATTCGGCGG.........CCCGCA 647
       :::|||         :::|||      ||||||         ||||||
 325 lyArgArgValArgSerAspThrSerAlaArgArgSerArgCysProAla 341

 648 TCCGGCCGGTTTGAGTGGCACGCACATTCATTTCATTGAGCCGGTCGGTG 697
     |||:::   ::::::|||                        :::||||||||
 342 SerSerProIleArgTrp.....................ThrGlyArgCy 351

 698 CAAACAAAACCGTTTGGACCATCAATTATCA...AGATGTAATTGCCATC 744
     |::::::   |||                ||| |||:::::::|||
 351 sArgArgTrpArgArgProLeuGlyCysSerProArgAlaThrCys.... 366

 745 GGACGTTTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTGATTGCTTT 794
       |||      |||:::|||   :::                   |||
 367 ..ThrAlaArgCysGlyArgAspGly.................CysSer 376

 795 GGGTGGTTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTTGGGTG 844
     :::::::|||::::::                |||
 377 AlaPhePheGlyAsnProLeuHisArgSerLeu.................. 387

 845 CGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACGCAGACAACCGC 894
                                        |||   |||
 388 ......................................ArgGlyProT 391

 895 GTGATTTCCGGTTCGGTATTGAACGG...CGCGATTACACAAGGCGCGCA 941
               |||      :::|||   |||:::   |||||||||:::
 391 rpAlaAlaProPheArgAlaHisArgSerArgSerThrThrArgArgCys 407

 942 CGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCC 991
       :::   |||:::      ::::::              ||||||||
 408 AlaValArgGlySerSerArgHisAspArgThrAlaSerThrArgArgPr 424

 992 GCAGCAAAGAGCTGTTCGGCTGGGTTGCGCC.................. 1022
     |:::::::              |||||||||
 424 oHisLysProPro......LysGlyCysAlaThrAspIleHisSerGlyA 439

1023 ...................GCAGCCGGACAAATACTCCATCACGCGTAC 1052
                         |||::::::::::             |||
 439 rgTyrCysTrpProArgThrAlaSerSerArgAlaAlaSerGlyAlaSer 455

1053 GACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTTCACGACAGCCG 1102
```

```
                |||         ::::::::::::      |||:::::::   ||
 456 AlaLysArgThrArgLeuArgArgArgSerCysProValArgSerProAr 472

1103 TCAACGG..............TGG.......................... 1112
     |:::|||              |||
 472 gArgArgGlyThrArgAlaAlaTrpHisSerAlaCysGlySerSerSerA 489

1113 ..CGACCGCGCCATGGTGCCGATTGGTACTTACGAGCGCGTAATGCC... 1157
       ||||||      |||      |||   :::      |||   :::
 489 rgArgProSerSerGlyArgProTrpSerValProIleArgProSerSer 505

1157 ..................................................... 1157

 506 IleCysGlyArgAlaValGlyLeuThrSerProSerSerProLeuAsnAr 522

1158 .......GCTAGACAT......CCTGCCTACCCTGCTTTTGCGCGATTTA 1194
             ||||||:::      ||||||               |||   |
 522 gProPheAlaArgArgSerAlaProAlaSerThrProCysArgArgHisA 539

1195 ATCGTCGGCGATACCGACAGCGCGCAAGCATTGGGTTGCTTGGAATTGGA 1244
     |||||||||||||                                  |||
 539 snArgArgArgTyr..............................Gly 544

1245 CGAAGAAGACCT 1256
        |||||||||
 545 SerArgArgPro 548
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04998

seq_documentation_block:

```
ID   AAY04998 standard; Protein; 388 AA.
XX
AC   AAY04998;
XX
DT   06-JUL-1999  (first entry)
XX
DE   Mycobacterium species protein sequence 50B.
XX
KW   Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW   hybridisation; detection; vaccine; immunisation; infection.
XX
OS   Mycobacterium sp.
XX
PN   WO9909186-A2.
XX
PD   25-FEB-1999.
XX
PF   14-AUG-1998;   98WO-FR01813.
XX
PR   11-SEP-1997;   97FR-0011325.
PR   14-AUG-1997;   97FR-0010404.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Gicquel B,  Lim EM,  Pelicic V,  Portnoi D,  Goguet de la Salmoniere Y;
PI   Guigueno A;
XX
DR   WPI; 1999-181045/15.
DR   N-PSDB; AAX34249.
XX
PT   Mycobacterial DNA vectors containing reporter constructs - for
PT   identifying coding or promoter sequences involved in
PT   infection-associated protein expression
XX
PS   Claim 32; Fig 50B; 309pp; French.
XX
CC   Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
CC   proteins from various Mycobacterium species microorganisms.  The
CC   encoding nucleotide sequences can be used as primers and probes for
CC   methods for detecting and identifying mycobacteria, especially belonging
CC   to the M. tuberculosis complex.  The encoded proteins can be used in
CC   vaccines for immunisation against a bacterial or viral infection.
XX
```

```
SQ   Sequence   388 AA;

alignment_scores:
           Quality:  131.00              Length:    355
             Ratio:  0.873                Gaps:     20
 Percent Similarity:  42.254   Percent Identity: 27.042

alignment_block:
 US-09-303-518D-127 x AAY04998   ..

 Align seg 1/1  to: AAY04998  from: 1  to: 388

     405 GTTCAGCAAAATCCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCA 454
         ::::::::::::   |||||||||   |||:::   |||      |||::
       1 IleArgArgGlyGlyCysArgArgArgArgGlyTrpArgSer...ArgAr 16

     455 ATGCGATGGACACCAATCCGCTNGCGGCAGACCCTGTGGTTGTGATCAAA 504
         :|||        :::   |||   |||                     |
      16 gCysGlyArgCysArgArgAlaValGly.....................A 26

     505 GAAGCCGNCGANGATTTCAGACGANGTNTGCTGGTATTGAGCCGTTTGAC 554
         ||   ||||||   :::|||
      26 rgArgArgArgArgTrpGln.............................. 32

     555 CGAGCGTAAAATCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCGTCTG 604
                                 |||||||:::|||
      33 ..............................TrpArgGluArg........ 36

     605 AAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCGGCC 654
          :::   |||||||||||::         |||   |||          |
      37 .ArgArgCysGlnHisArgGlnArgTrpArgArgTrpProArgArgCysA 53

     655 GGTTTGAGTGGCACGCACATTCATTTCATTGAGCCGGTCGGTGCAAACAA 704
         ||:::::::|||:::                |||   |||   ::::::
      53 rgTrpArgTrpArgMet..............AlaLeuArgArgArgArg 64

     705 AACCGTTTGGACCATCAATTATCAAGATGTAATTGCCATCGGACGTTTGT 754
         :::|||      ::::::      |||
      65 GlyArgArgThrArgArgThrArgArg..................... 73

     755 TTGCAACAGGCCGTCTGAACACCGAGCGCGT.................. 785
             |||||||||   ::::::|||   |||
      74 ....AsnArgProArgArgArgArgArgArgArgGlyProGlyArgA 89

     786 ........GATTGCTTTGGGTGGTTCTCAAGTCAACAAACCACG...... 821
                       |||::::::      :::::::|||
      89 rgArgProArgThrValGlyTyrTrpArgArgArgArgThrArgArgAla 105

     822 ..................CCTCTTGCGTACCGTTTTGGGTGCGAAAGTAT 853
                          |||   |||   |||   |||   :::
     106 ArgArgTrpTyrArgGlyProThrAlaAlaArgSerGlyArgHis..... 120

     854 CGCAAATTACTGCGGGCGAATTGGTTGACGCAGACAACCGCGTGATTTCC 903
                   ||||||   |||   ||||||                |
     121 .............GlyArgArgGlyTrpArgArgTrpAlaAspArgGlnA 133

     904 GGTTCGGTATTGAACGGCGC.......................GATTAC 929
         ||   |||   :::||||||
     133 rgArgGlyArgArgArgArgArgCysArgArgValArgArgGlyArgArg 149

     930 ACAAGGCGCGCACGATTATTTGGGACGCTACCACAATCAGATTTCCGTTA 979
         :::||||||:::|||                ::::::      |||
     150 SerArgArgCysArgArgGluArgHisAlaAspArgAlaArgArgArgAr 166

     980 TCGAAGAAGGCCGCAGCAA..............AGAGCTGTTCGGCTGG 1014
          |||||||||   :::|||               |||:::   |||   |
     166 gArgArgArgArgArgGlnGlnPheArg***TrpArgGlyArgArgArgG 183

    1015 GT...TGCGCCGCAGCCGGACAAATACTCCATCACGCGTACGACCCTCGG 1061
         ||   |||      |||                :::::      ||||||
     183 lyArgCysArgArgAlaProLeuArgGlnTrpArgValArgArgProArg 199

    1062 CCATTTCCTGAAAAACAAACTCTTCAAGTTCACGACAGCCGTCAACGGTG 1111
                       ::::::   :::   ::::::   |||   |||||
      200 ..............ArgSerArgHisGlyArgGlnHisArgTyrArgTr 211

    1112 GCGACCGCGC............CATGGTGCCGAT.........TGGTACT 1140
         ||||   |||            :::   ||||||          |||
      211 pArgArgArgArgArgTrpGlyArgGlnAlaAspArgProArgTrpArgA 228

    1141 TACGAGCGCGTAATGCCGCTAGACATCCT...GCCTACCCTGCT...TTT 1184
           |||:::|||::::::   |||:::      ::::::|||:::
      228 rgArgCysArgArgGlyProArgArgSerLeuGlyTrpPro***ArgArg 244

    1185 GCGCGATTTAATCGTCGGCGATAC............CGACAGCGCGCAAG 1222
         ||||||::::::|||   |||:::            |||||||||
      245 AlaArgTrpGluArgGlyArgTrpArgProAlaIleArgGlnArgArg.. 260

    1223 CATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTC 1272
                               ||||||||||||||||      :::
      261 .....................ArgArgArgProArgHisArgArgAsnT 270

    1273 GTCTGCCCGGGCAAATACGAATANGGCCCGCTGTTGCGTAAGGTGCTGGA 1322
               |||      ||||||      :::     |||||||||  :
      270 hrAlaGlyGlyGlyGluArgIleGlyAspGlyPhe..ValArgCysThrA 286

    1323 AACCNTTGAGAAGG 1336
         ::|||   ||||||
      286 rgProThrArgArg 290

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59827

seq_documentation_block:
ID   AAB59827 standard; Protein; 1592 AA.
XX
AC   AAB59827;
XX
DT   04-APR-2001  (first entry)
XX
DE   Protein #4 encoded by TutD/E gene.
XX
KW   Toluene degradation; enzyme; waste degradation; TutE; TutD.
XX
OS   Thauera aromatica.
OS   Xanthomonas maltophilia.
OS   Geobacter metallireducens.
OS   Azoarcus tolulyticus.
XX
PN   WO200072650-A2.
XX
PD   07-DEC-2000.
XX
PF   24-MAY-2000; 2000WO-US14298.
XX
PR   01-JUN-1999;   99US-0323872.
XX
PA   (UYOH-) UNIV OHIO.
XX
PI   Coschigano PW;
XX
DR   WPI; 2001-041080/05.
DR   N-PSDB; AAF23627.
XX
PT   Composition comprising toluene degrading enzyme useful for biological
PT   treatment of organic compounds, especially for degrading toluene or its
PT   analogs  -
XX
PS   Disclosure; Fig 12; 122pp; English.
XX
CC   The present invention relates to toluene degrading enzyme genes and
CC   proteins tutH (see AAF23629 and AAB59831), tutI (AAF23630 and AAB59832),
CC   tutF (AAF23631 and AAB59833) and tutG (AAF23632 and AAB59834). The
CC   toluene degrading enzymes are homologues of pyruvate formate lyase. The
CC   toluene degrading enzymes are useful for biological treatment of organic
CC   compounds and in particular for the degradation of toluene and its
```

```
CC   analogs contained in liquid or solid waste source. The present sequence
CC   is a protein sequence encoded by toluene degrading enzyme gene, TutD/E.
XX
SQ   Sequence   1592 AA;


alignment_scores:
           Quality:  128.00              Length:    504
             Ratio:  0.670                Gaps:     26
 Percent Similarity:  37.897   Percent Identity: 22.222

alignment_block:
 US-09-303-518D-127 x AAB59827   ..

 Align seg 1/1  to: AAB59827  from: 1  to: 1592

      81 CGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGCCCCTNGATGA 130
         |||||||||   :::      |||   ||||||:::::      :::::
     712 ArgSerArgCysSerProAspArgCysCysArgTrpSerArgCysSerSe 728

     131 AAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGAC 180
         :   :::   |||||||||      :::   |||::::::
     728 rProSerProArgArgCysProProSerSerProAlaGly..........741

     181 AAAAAGNATCCGGGCGTGGTGTTTACCGCGCCNGTTTCAGGCAAAATCGC 230
           :::   |||           |||      |||   :::::::|||
     742 .....AlaProGlyAlaThrCysSerArgArgProPheSerArgSerArg 756

     231 CGCCATCCATCGCGGCGAAAAGCGCGTACTTCAGTCGGTCGTGATTGCCG 280
               :::   |||   ||||||           |||   |||||
     757 AspSerAlaGlyProArgAlaAlaArgPhe.........ArgArgCysAr 770

     281 TTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGCGCCCGAAGCGTTG 330
         |         :::|||   |||:::|||           ||||||:::
     770 gAspAlaCysGluArg...ArgAlaArgCysProGlyProArgSerAlaP 786

     331 GCAAACTTAAGCGG.................CGANGAANTNNGNNGCAA 362
           ::::::::::|||                |||::::::   ::::::
     786 roSerIleArgArgGlySerArgAspArgSerArgAlaSerArgSerArg 802

     363 TCTGATCCAATCCGGTTTGTGGACTGCGCTGCGTAN.............. 398
         ::::::|||:::            |||:::|||:::
     803 GlySerProLeu............CysGlyAlaThrAlaThrSerCysPr 815

     399 ..............................................CCGT 402
                                                      |||
     815 oArgArgArgArgCysSerIleGlyAlaSerSerGlyCysProHisProP 832

     403 CCGTTCAGCAAAATCCCTGC.....................CGTCGATGC 431
           |||:::::::::|||                       |||||||||
     832 roValArgArgSerProValAsnSerSerLysArgAlaHisArgArgCys 848

     432 CGAGCCGTTCGC................................CATCT 448
            |||   |||
     849 ThrAlaArgArgGlyArgPheArgGlyProThrSerArgAspThrGlyAr 865

     449 TCGTCAATGCGATGGACACCAATCCGCTNGCGGCAGACCCTGTGGTTGTG 498
          |||:::|||         :::          |||   |||   |||:
     865 gArgArgCys......TrpArgTrpProArgProArgArgCysArgCysS 880

     499 ATCAAAGAAGCCGNCGANGATTTCAGACGANGTNTGCTGGTATTGAGCCG 548
         :::::|||      |||   :::   :::
     880 erArgArgTrpGlyArgProLeuTrpAla..................... 889

     549 TTTGACCGAGCGTAAAATCCATGTGTGTAAGGCAGCTGGCGCAGA..... 593
                       :::   |||      :::   |||||||||
     890 ..............SerGlyCysProArgAlaArgTrpArgArgGlySe 901

     594 ............................CGTGCCGTCTGAAAATGCTGCC 615
                                        :::      :::   |||
     901 rAsnTrpSerSerGlyArgSerSerAlaAlaSerProLysArgThrCysG 918

     616 AACATCGA.........AACACATGAATTCGGCGG.........CCCGCA 647

         :::|||         :::|||      ||||||         ||||||
     918 lyArgArgValArgSerAspThrSerAlaArgArgSerArgCysProAla 934

     648 TCCGGCCGGTTTGAGTGGCACGCACATTCATTTCATTGAGCCGGTCGGTG 697
         |||:::   ::::::|||                      :::|||||||||
     935 SerSerProIleArgTrp.....................ThrGlyArgCy 944

     698 CAAACAAAACCGTTTGGACCATCAATTATCA...AGATGTAATTGCCATC 744
         |:::::::   |||                |||   |||::::::|||
     944 sArgArgTrpArgArgProLeuGlyCysSerProArgAlaThrCys.... 959

     745 GGACGTTTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGTGATTGCTTT 794
           |||      |||:::|||   :::                  |||
     960 ..ThrAlaArgCysGlyArgAspGly.................CysSer 969

     795 GGGTGGTTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTTGGGTG 844
         ::::::|||::::::              |||
     970 AlaPhePheGlyAsnProLeuHisArgSerLeu..................... 980

     845 CGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACGCAGACAACCGC 894
                                                  |||   |||
     981 .....................................ArgGlyProT 984

     895 GTGATTTCCGGTTCGGTATTGAACGG...CGCGATTACACAAGGCGCGCA 941
                   |||      :::|||   |||:::   |||||||||:::
     984 rpAlaAlaProPheArgAlaHisArgSerArgSerThrThrArgArgCys 1000

     942 CGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCC 991
           :::   |||:::      ::::::                ||||||||
    1001 AlaValArgGlySerSerArgHisAspArgThrAlaSerThrArgArgPr 1017

     992 GCAGCAAAGAGCTGTTCGGCTGGGTTGCGCC..................... 1022
         |::::::               |||||||||
    1017 oHisLysProPro......LysGlyCysAlaThrAspIleHisSerGlyA 1032

    1023 ...................GCAGCCGGACAAATACTCCATCACGCGTAC 1052
                             |||:::::::::            |||
    1032 rgTyrCysTrpProArgThrAlaSerSerArgAlaAlaSerGlyAlaSer 1048

    1053 GACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTTCACGACAGCCG 1102
           |||          ::::::::::::    |||::::::   ||
    1049 AlaLysArgThrArgLeuArgArgArgSerCysProValArgSerProAr 1065

    1103 TCAACGG..............TGG........................... 1112
         |:::|||              |||
    1065 gArgArgGlyThrArgAlaAlaTrpHisSerAlaCysGlySerSerSerA 1082

    1113 ..CGACCGCGCCATGGTGCCGATTGGTACTTACGAGCGCGTAATGCC... 1157
           ||||||      |||      |||   :::      |||   :::
    1082 rgArgProSerSerGlyArgProTrpSerValProIleArgProSerSer 1098

    1157 ..................................................... 1157

    1099 IleCysGlyArgAlaValGlyLeuThrSerProSerSerProLeuAsnAr 1115

    1158 .......GCTAGACAT......CCTGCCTACCCTGCTTTTGCGCGATTTA 1194
                 ||||||:::      ||||||                |||   |
    1115 gProPheAlaArgArgSerAlaProAlaSerThrProCysArgArgHisA 1132

    1195 ATCGTCGGCGATACCGACAGCGCGCAAGCATTGGGTTGCTTGGAATTGGA 1244
         ||||||||||||||                                 |||
    1132 snArgArgArgTyr................................Gly 1137

    1245 CGAAGAAGACCT 1256
            |||||||||
    1138 SerArgArgPro 1141

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB27242

seq_documentation_block:
ID   AAB27242 standard; Protein; 571 AA.
XX
AC   AAB27242;
```

```
XX
DT   27-MAR-2001  (first entry)
XX
DE   Human EXMAD-20 SEQ ID NO: 20.
XX
KW   Extracellular matrix and adhesion-associated protein; EXMAD; cancer;
KW   inflammation; reproductive disorder; cardiovascular disorder;
KW   immune disorder; musculoskeletal disorder; developmental disorder;
KW   gastrointestinal disorder; cell proliferation disorder.
XX
OS   Homo sapiens.
XX
PN   WO200068380-A2.
XX
PD   16-NOV-2000.
XX
PF   10-MAY-2000; 2000WO-US12811.
XX
PR   11-MAY-1999;   99US-0133643.
PR   23-AUG-1999;   99US-0150409.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Bandman O,  Hillman JL,  Tang YT,  Lal P,  Yue H,  Baughn MR,  Lu DAM;
PI   Azimzai Y;
XX
DR   WPI; 2001-007395/01.
DR   N-PSDB; AAC66909.
XX
PT   Isolated polynucleotide encoding extracellular matrix or
PT   adhesion-associated protein (EXMAD) useful for diagnosing, treating, or
PT   preventing disorders associated with expression of EXMAD such as
PT   proliferative, immune and genetic disorders -
XX
PS   Claim 1; Page 106-107; 129pp; English.
XX
CC   The present invention provides the protein and coding sequences for 25
CC   novel extracellular matrix and adhesion-associated proteins (EXMADs).
CC   These are designated EXMAD-1, EXMAD-2, EXMAD-3, EXMAD-4, EXMAD-5,
CC   EXMAD-6, EXMAD-7, EXMAD-8, EXMAD-9, EXMAD-10, EXMAD-11, EXMAD-12,
CC   EXMAD-13, EXMAD-14, EXMAD-15, EXMAD-16, EXMAD-17, EXMAD-18, EXMAD-19,
CC   EXMAD-20, EXMAD-21,  EXMAD-22, EXMAD-23, EXMAD-24 and EXMAD-25. They are
CC   useful in the prevention and treatment of cancers, cell proliferation,
CC   cardiovascular, reproductive, immune, musculoskeletal, developmental and
CC   gastrointestinal disorders and inflammation.
XX
SQ   Sequence   571 AA;


alignment_scores:
            Quality:  122.50              Length:    462
              Ratio:  0.554                Gaps:     20
 Percent Similarity:  47.835   Percent Identity: 20.346

alignment_block:
 US-09-303-518D-127/rev x AAB27242   ..

 Align seg 1/1  to: AAB27242  from: 1  to: 571

    1340 CCTTCCTTCTCAANGGTTTCCAGCACCTTACGCAACAGCGGGCCNTATTC 1291
         |||   :::   ::::::|||:::|||   :::   |||:::   :::::
      23 ProMetMetProThrThrSerGlyThrSerGlnAlaSerSerSerPheAs 39

    1290 GTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCA 1241
         :           :::|||:::||||||         |||   :::|||:
      39 nThrAlaLysThrSerThrSerLeuHis.........SerHisThrSerS 53

    1240 ATTCCAAGCAACCCAATGCTTGCGCGCTGTCGGTA.............. 1206
         :::::   :::|||::::::         |||:::
      53 erThrHisHisProGluValThrProThrSerIleThrAsnIleThrLeu 69

    1205 .......................TCGCCGACGATTAAATCGCGCAAAAG 1180
                               :::|||:::      :::   :::::

      70 AsnProThrSerIleGlyThrTrpThrProValAlaHisThrThrSerAl 86

    1179 CAGGGTAGGCAGGATGTCTAGCGGCATTACGCGCTCGTAAGTACCAATCG 1130
         :      :::|||:::::::::   :::|||            |||   |
      86 aThrSerSerArgLeuThrThrProPheThrThrHisSerProProThrG 103

    1129 GCACC......ATGGCGCGGTCGCCACCGTTGACGGCTGTCGTGAACTTG 1086
         ||:::      ::::::   :::   |||:::|||||||:::
     103 lySerSerProIleSerSerThrGlyProMetThrAlaThrSerPheGln 119

    1085 AAGAGTTTGTTTTTCAGGAAATGGCCGAGGGTCGTACGCGTGATGGAGTA 1036
         :::   ::::::
     120 ThrThrThrTyrTyrThrPro..............................  126

    1035 TTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCCTT 986
                        :::::::|||:::::::::|||
     127 ................ProSerHisProGlnThrThrLeu.........  134

     985 CTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCGTGCGCG 936

     134 ..................................................  134

     935 CCTTGTGTAATCGCGCCGTTCAATACCGAA.........CCGGAAATC..  897
         |||       :::   |||||||:::|||:::         |||:::
     135 ProThrHisValProProPheSerThrSerLeuValThrProSerThrHi 151

     896 .....................ACGCGGTTGTCTGCGTCAACCAATTCGC 869
                               |||:::::::::::::|||::::::
     151 sThrValIleIleThrThrHisThrGlnMetAlaThrSerAlaSerIleH 168

     868 CCGCAGTAATTTGCGATACTTTCGCACCCAAAACGGTACGCAAGAGGCGT 819
           ::::::          |||      |||   |||:::   |||
     168 isSerThrProThrGlyThrValProProProThrThrLeuLysAlaThr 184

     818 GGTTTGTTGACTTGAGAACCACCCAAAGCAATC................... 786
         |||                |||||   ::::::
     185 GlySerThrHisThrAlaProProMetThrValThrThrSerGlyThrSe 201

     785 ....ACGCGCTCGGTGTTCAGACGGCCTGTTGCAAACAAACGTCCGATGG 740
             |||:::|||   |||         :::|||::::::      ::::
     201 rGlnThrHisSerSerPheSerThrAlaThrAlaSerSerSerPheIleS 218

     739 CAATTACATCT...TGATAATTGATGGTCCAAACGGTTTTGTTTGCACCG 693
         ::   :::|||            :::   |||:::            |||
     218 erSerSerSerTrpSerSerTrpLeuProGlnAsnSerSerSerArgPro 234

     692 ACCGGCTCAATGAAATGAATGTGCGTGCCACTCAAACCGGCCGGATGCGG 643
            :::|||               :::|||         ::::::
     235 ProSerSerProIleThrThrGlnLeuProHisLeuSerSerAlaThrTh 251

     642 GCCGCCGAATTCATGTGTTTCGATGTTGGCAGCATTTTCAGACGGCACGT 593
          |||   ::::::      ::::::   ::::::||||||   :::   |
     251 rProValSerThrThrAsnGlnLeuSerSerSerPheSerProSerProS 268

     592 CTGCGCCAGCTGCCTTACACACATGGATTTTACGCTCGGTCAAACGGCTC 543
         ||||||||:::::::::   :::::::::
     268 erAlaProSerThrValSerSerTyrVal..................... 277

     542 AATACCAGCANACNTCGTCTGAAATCNTCGNCGGCTTCTTTGATCACAAC 493
            :::|||   :::      ::::::|||:::   |||:::   |||::
     278 ProSerSerHisSerSerProGlnThrSerSerProSerValGlyThrSe 294

     492 CACAGGGTCTGCCGCNAGCGGATTGGTGTCCATCGCATTGACGAAGATGG 443
         :::::::   :::::::::   :::   |||   :::|||::::::   :
     294 rSerSerPheValSerAlaProValHisSerThrThrLeuSerSerGlyS 311

     442 CGAACGGCTCGGCATCGACGGCAGGGATTTTGCTGAACGGACGGNTACGC 393
         :::::::::|||   ||||||
     311 erHisSerSerLeuSerThr.............................. 317

     392 AGCGCAGTCCACAAACCGGATTGGATCAGATTGCNNCNNANTTCNTCGCC 343
                |||                      :::   ::::::|||||
     318 .........HisPro.........ThrThrAlaSerValSerAlaSerPr 328
```

```
     342 GCTTAAGTTTGCCAACGCTTCGGGCGCGTAGCGTTCGAACTCGATT...T 296
         ||||  |||  ::::::  :::|||     |||::::::|||  :
     328 oLeu...PheProSerSerProAlaAla......SerThrThrIleArgA 342

     295 CGTCGTTGCCTTCAACGGCAATCACGACCGACTGAAGTACGCGCTTTTCG 246
         :::::||||||  |||                              |||.
     342 laThrLeuProHisThrIle.....................SerSer 350

     245 CCGCGATGGATGGCGGCGATTTTGCCT...GAAACNGGCGCGGTAAACAC 199
         |||      ::::::|||:::||||||  :::|||  :::|||:::
     351 ProPheThrLeuSerAlaLeuLeuProIleSerThrValThrValSerPr 367

     198 CACGCCCGGATNCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGACGG 149
          ||||||::::::  |||:::|||  ||||||     |||     :
     367 oThrProSerSerHisLeuAlaSer...SerThrIleAlaPhe...ProS 382

     148 CATCGCCTTCCTTGACTTTCATCNAGGGGCGCATACCGGCATATTCTTCG 99
         :::::|||      |||      :::         ||||||:::||||||
     382 erThrProArgThrThrAlaSerThrHisThrAlaProAlaPheSerSer 398

      98 .........CCAAGCAACGCGACTTCGGTAATGACG 72
                 |||::::::|||||:::  |||
     399 GlnSerThrThrSerArgSerThrSerLeuThrThr 410

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB94220

seq_documentation_block:
ID   AAB94220 standard; Protein; 639 AA.
XX
AC   AAB94220;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:14581.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 14581; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   639 AA;


alignment_scores:
            Quality:  122.00              Length:    496
              Ratio:  0.545                 Gaps:     22
 Percent Similarity:  45.161   Percent Identity: 22.581

alignment_block:
 US-09-303-518D-127/rev x AAB94220   ..

 Align seg 1/1  to: AAB94220  from: 1  to: 639

    1304 AGCGGGCCNTATTCGTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAG 1255
         |||:::|||::::::   ||||||   :::|||:::
      95 SerAlaProPheThrSerLeuProPheSerThrSer.............. 106

    1254 GTCTTCTTCGTCCAATTCCAAGCAACCCAATGCTTGCGCGCTGTCG.... 1209
          ||||||:::::::::::::::::::||||||::::::::::||||||
     107 .SerSerAlaAlaSerThrSerAsnProAsnSerAlaSerLeuSerServ 123

    1208 ......................GTATCGCCGACGATTAAATCGCGCAAA 1182
                                :::   ||||||  ::::::   :::
     123 alPheAlaGlyLeuProLeuProLeuProProThrSerGlnGlyLeuSer 139

    1181 AGCAGGGTAGGCAGGATGTCTAGCGGCATTACGCGCTCGTAAGTA...CC 1135
         :::   :::      ::::::::::|||   |||   |||   :::   ||
     140 AsnProThrProValIleAlaGlyGlySerThrProSerValAlaGlyPr 156

    1134 AATCGGCACCATGGCGCGGTCGCCACCGTTGACGGCT............. 1098
         |:::|||      ::::::||||||   |||:::|||
     156 oLeuGly......ValAsnSerProLeuLeuSerAlaLeuLysGlyPheL 171

    1098 ..................................................... 1098

     171 euThrSerAsnAspThrAsnLeuIleAsnSerSerAlaLeuSerSerAla 187

    1097 ............GTCGTGAACTTGAAGAGTTTGTTTTTCAGGAAATGGCC 1060
                     ::::::::::|||:::||||||   :::::::::
     188 ValThrSerGlyLeuAlaSerLeuSerSerLeuThrLeuGlnAsn..... 202

    1059 GAGGGTCGTACGCGTGATGGAGTATTTGTCCGGCTGCGGCGCAACCCAGC 1010
                                      |||      :::|||:::   |
     203 ............................SerAspSerSerAlaSerAlaP 210

    1009 CGAACAGCTCTTTGCTGCGGCCTTCTTCGATAACGGAAATCTGATTGTGG 960
         |||||   :::          ||||||:::|||
     210 roAsn...LysCysTyrAlaProSerAlaIlePro.............. 220

     959 TAGCGTCCCAAATAATCGTGCGCGCCTTGTGTAATCGCGCCGTTCAATAC 910
             |||:::   :::   :::|||      :::|||   |||
     221 ...ThrProGlnArgThrSerThrPro...GlyLeuAlaLeuPheProGl 235

     909 CGAACCGGAAATCACGCGGTTGTCTGCGTCAACC................ 876
             |||:::   :::      |||:::||||||
     235 yProProSerProValAlaAsnSerThrSerThrProLeuThrLeuProV 252
```

```
875 ..AATTCGCCC.........GCAGTAATTTGCGATACTTTCGCACCCAAA 837
    :::||||||          |||:::   ::::::|||   ||||||
252 alGlnSerProLeuAlaThrAlaAlaSerAlaSerThrSerAlaProVal 268

836 ACGGTACGCAAGAGGCGTGGTTTGTTGACTTGAGAACCACCCAAAGCAAT 787
    :::      :::      :::||||||            |||   :::
269 SerCysGlySerSerAlaSerLeuLeuArgGlyProHisProGlyThrSe 285

786 CACGCGCTCGGTGTTCAGACGGCCTGTTGCAAACAAACGTCCGATGGCAA 737
              :::         |||:::|||:::      |||
285 rAspLeuHisIleSerSerThrProAlaAlaThrThrLeuPro....... 299

736 TTACATCTTGATAATTGATGGTCCAAACGGTTTTGTTTGCACCGACCGGC 687
                  :::|||::::::|||         :::|||||
300 .............ValMetIleLysThrGluProThrSerProThrPro 311

686 TCAATGAAATGAATGTGCGTGCCACTCAAA...CCGGCCGGATGCGGGCC 640
    |||                    :::|||   |||:::      |||
312 Ser.................AlaPheLysGlyProSerHisSerGlyAs 322

639 GCCGAAT............TCATGTGTTTCGATGTTGGCAGCATTTTCAG 602
     |||:::            :::      :::   |||:::
322 nProSerHisGlyThrLeuGlyLeuSerGlyThrLeuGlyArgAlaTyrT 339

601 ACGGCACGTCTGCGCCAGCTGCCTTACACACATGGATTTTACGCTCGGTC 552
      :::||||||:::|||   :::|||   :::
339 hrSerThrSerValProIleSerLeuSerAlaCys.............. 350

551 AAACGGCTCAATACCAGCANACNTCGTCTGAAATCNTCGNCGGCTTCTTT 502
        ||||||   :::   :::   |||:::||||||:::::   ||
351 ......LeuAsnProAlaLeuSerGlyLeuSerSerSerSerThrProLe 365

501 GATCACAACCACAGGGTCTGCCGCNAGCGGATTGGTGTCCATCGCATTGA 452
    |
365 u..................................................365

451 CGAAGATGGCGAACGGCTCGGCATCGACGGCAGGGATTTTGCTGAACGGA 402
              |||||||||          :::::::|||   |||
366 ...........AsnGlySerAsnProLeuSerSerIleSerLeuProPro 378

401 CGGNTACGCAGCGCAGTCCACAAACCGGATTGGATCAGATTGCNNCNNAN 352
    :::      |||:::          |||   :::      |||   :::
379 HisGlySerSerThrProIleAlaProValPheThrAlaLeuProSerPh 395

351 TTCNTCGCCGCTTAAGTTTGCCAACGCTTCGGGCGCGTAGCGTTCGAACT 302
     :::|||      :::          :::|||              |||
395 eThrSerLeuThrAsnAsnPheProLeuThrGlyAsnProSerLeuAsnP 412

301 CGATTTCGTCGTTGCCTTCAACGGCAATCACGACCGACTGAAGTACGCGC 252
            |||||||||::::::   |||:::|||:::   ::::::
412 roSerValSerLeuProGlySerLeuIleAlaThrSerSerThrAlaAla 428

251 TTTTCG..............CCGCGATGGATGGCGGCGATTTTG..... 222
    |||                |||         ||||||:::|||
429 ThrSerThrSerLeuProHisProSerSerThrAlaAlaValLeuSerGl 445

221 ..............CCTGAAACNGGCGCG.................... 207
                  |||   ::::::|||
445 yLeuSerAlaSerAlaProValSerAlaAlaProPheProLeuAsnLeuS 462

206 ..................GTAAACACCACGCCCGGATNCTTTTTGTCT 177
                       ::::::|||   |||   :::   |||
462 erThrAlaValProSerLeuPheSerValThrGlnGlyProLeuSerSer 478

176 TCAAACAGCACTTGGCCTTTTTTGACGGCATCG......CCTTCCTTGAC 133
    ||||||   ::::::|||   :::::::::|||      ||||||:::||
479 SerAsnLeuSerTyrProGlyPheSerValSerAsnThrProSerValTh 495

132 TTTCATCNAGGGGCGCATACCGGCATATTCTTCGCCAAGCAACGCGACTT 83
    |               :::|||:::   ::::::||||||::::::::::
495 rProAlaLeuProSerPheProGlyLeuGlnAlaProSerThrValAlaA 512

 82 CGGTA.................ATGACGGGCCCGTCA 63
```

```
       ::|||                        |||:::||||||
512 laValThrProLeuProValAlaAlaThrAlaProSer 524
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG03731

seq_documentation_block:

ID ABG03731 standard; Protein; 696 AA.
XX
AC ABG03731;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #3722.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS67918.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 34090; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 696 AA;

alignment_scores:

Quality:	121.00	Length:	467
Ratio:	0.571	Gaps:	23
Percent Similarity:	45.396	Percent Identity:	23.555

alignment_block:

```
US-09-303-518D-127 x ABG03731   ..

Align seg 1/1  to: ABG03731  from: 1  to: 696

      9 AATCAAAAAAGGTCTAAACCTGCCCATCGCGGGCAGACCGGAGCAAGTCA 58
        |||   :::      |||||||:::   |||||||:::      :::   ::
     14 AsnLeuArgAlaArgLysProSerLysArgGlyLysAspArg***ArgAr 30

     59 TTTATGACGGGCCCCGTCATTACCGAAGTCGCGTTGCTTGGCGAAGAATAT 108
        ::::      :::|||      :::   |||   |||
     30 gValSerProThrArgSerGlyLysArgArgGlyAla..... 42

    109 GCCGGTATGCGCCCCTNGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAA 158
                          |||:::::::::|||:::   ::::::|||
     43 ...................GluGluLysAsnArgGlnGluLysLysLys 52

    159 AGGCCAAGTGCTGTTTGAAGACAAAAAGNATCCGGGCGTGGTGTTTACCG 208
            :::          |||:::|||   |||   |||          ||
     53 GlyArgGluLysGluArgArgGluLysArgSerGluArgGlnArgAspAr 69

    209 CGCCNGTTTCAGGCAAAATCGCCGCCATCCATCGCGGCGAAAAGCGCGTA 258
        |         ::::::::::|||:::::::   :::|||||||:::||||||
     69 gArgArgArgLysGluGlnArgLysGluGluGlnArgArgArgAlaArgT 86

    259 CTTCAGTCGGTCGTGATTGCCGTTGAAGGCAACGACGAAATCGAGTTCGA 308
        ||:::                          :::|||:::   |||
     86 hrAsn.......................GluArgLysProArg...... 92

    309 ACGCTACGCGCCCGAAGCGTTGGCAAACTTAAGCGGCGANGAANTNNGNN 358
              ::::::::::::   |||       :::          ::::::::
     93 ......GlnThrGlnAlaAsnGlyAlaThrSerSer***LysAlaSerAl 107

    359 GCAATCTGATCCAATCCGGTTTGTGGACTGCGCTGCGTANCCGTCCGTTC 408
        :|||:::      :::                 ::::::::::|||:::::::
    107 aGlnGlnAlaGlyMet............TrpGlyGlySerPro***ThrA 120

    409 AGCAAAATCCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGC 458
        ::   :::      ||||||      |||           |||:::
    120 spAlaThrAlaIleArgArgGlyGlyAlaProCysSerSerArgArgThr 136

    459 GATGGACACCAATCCGCTNGCGGCAGACCCTGTGGTTGTGATCAAAGAAG 508
              :::|||::::::   :::   |||   |||
    137 CysLeuAsnGlnGlyThrIleAlaThrProSerGly............Ar 149

    509 CCGNCGANGATTTCAGACGANGTNTGCTGGTATTGAGCCGTTTGACCGAG 558
         ||||||      ::::::         ||| |||:::         :::
    149 gArgArgHisGlyAspAlaGly***ProGly.LeuAlaSerGluHisAsp 165

    559 CGTAAAATCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAA 608
           :::   |||   |||      :::||||||      ||||||:::::
    166 AlaSerGlyHisGlyCysLeuArgThrGlyAlaGly***ProSerAspSe 182

    609 T.........................GCTGCCAACATCGAAACACATG 631
        :                         |||   ::::::   |||||||
    182 rThrGluSerValCysArgArgProLeuAlaMetHisValProThrHisG 199

    632 AATTCGGCGGCCCGCATCCGGCCGGTTTGAGTGGCACGCACATTC..... 676
        ||      ||||||      :::   ||| |||::::::|||||||
    199 luSerHisGlyProValPheThrArgLeu.ValSerHisThrPheHisCy 215

    677 .............ATTTCATTGAGCCGGTCGGTGCAAACAAAACCG...T 710
                      ::::::::::|||   |||      :::|||   :
    215 sGlySerLysLeuProAlaValGlyArgProValAlaCysArgProThrT 232

    711 TTGGACCATCAATTATCAAGATGTAATTGCCATCGGACGTTTGTTTGCAA 760
        :::::||||||  |||             ||||||:::           ::
    232 yrSerProSer.Leu............CysHisAsnPro.........Gl 241

    761 CAGGCCGTCTGAACACCGAGCGCGTGATTGCTTTGGGTGGTTCTCAAGTC 810
        :||||||::::::      |||::::::              |||:::
    241 nArgProAlaGlnLeuLeuAlaHisSer..............SerAlaL 253

    811 AACAAACCACGCCTCTTGCGTACCGTTTTGGGTGCGAAAGTATCGCAAAT 860
```

```
          |||   :::||||||::::::                          :::
    253 euGlnCysAlaProLeuSerTrpAsp.................ProGln 263

    861 TACTGCGGGCGAATTGGTTGACGCAGACAACCGCGTGATTTCCGGTTCGG 910
          |||:::                  :::||||||      |||   ||
    264 ArgCysAla............ProProSerProArgProHisArgArgGl 276

    911 TATTGAACGGCGCGATTACACAAGGCGCGCACGATTATTTGGGACGCTAC 960
        |                       |||||||||                 |
    276 yProProSerProHisProHisArgArgAla.................P 287

    961 CACAATCAGATTTCCGTTATCGA........................... 983
        ||   |||          :::|||
    287 roProSerProHisProHisArgArgAlaHisThrThrAlaArgThrAsp 303

    984 .......................AGAAGGCCGCAGCAAAGAGCTGTTCG 1009
                                |||:::   |||:::||||||:::||
    304 ProThrThrSerAlaProProProArgGlnThrGlnArgArgAlaThrAr 320

   1010 GCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGTACGACCCTC 1059
        |      ::::::         :::      |||            ||||
    320 gGluProAlaThrLysHisThrArgAsnAlaHis............ProA 333

   1060 GGCCATTTCCTGAAAAACAAACTCTTCAAGTTCACGACAGCCGTCAACGG 1109
        ||              ::::::      ::::::|||      |||:::|||
    333 rgArgSerAlaCysAsnArgGlyThrHisThrHisProArgArgArgArg 349

   1110 TGG.......................CGACCGCGCCAT......GGTGC 1129
                                  |||||||||:::         ||
    350 ThrThrGluArgThrThrHisHisAlaArgProArgAsnArgGlyGlnAl 366

   1130 CGATTGGTACTTACGAGCGCGTAATGCCGCTAGACATCCTGCCTACCCTG 1179
        |            |||      :::|||:::||||||            :
    366 aThrProAsnThrArgGlnProThrAlaGlyArgHisGluGluThrAspG 383

   1180 CTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAAGCATTGGG 1229
        ::   :::|||      ||||||::::::   |||      :::   |||
    383 lyAlaThrArg......ArgArgGlnHisGlyGlnThrArgGlyGluGly 397

   1230 TTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCGT 1274
                      ||||||||||||   |||   ::::::   |||
    398 ............GlyArgArgArgGlyArgAlaAlaLysThrArg 408

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23389

seq_documentation_block:
ID   ABG23389 standard; Protein; 1194 AA.
XX
AC   ABG23389;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #23380.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
```

```
DR   N-PSDB; AAS87576.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 53748; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1194 AA;


alignment_scores:
            Quality:  113.50              Length:    441
              Ratio:  0.565                Gaps:     18
 Percent Similarity:  45.578   Percent Identity: 21.995

alignment_block:
 US-09-303-518D-127/rev x ABG23389   ..

 Align seg 1/1  to: ABG23389  from: 1  to: 1194

    1313 TTACGCAACAGCGGGCCNTATTCGTATTTGCCCGGGCAGACGAAGCTGCA 1264
          |||   :::|||||||||   :::   ||||||
     584 LeuSerGlnSerGlyProProGlyLeuLeuPro..... 594

    1263 CAAAGCGAGGTCTTCTTCGTCCAATTCCAAGCAACCCAATGCT...TGCG 1217
                 |||   |||   :::|||||||   |||::::::     :
     595 .........SerProSerPheAspSerLysProProThrThrLeuLeuG 608

    1216 CGCTGTCGGTATCGCCGACGATTAAATCGCGCAAAAGCAGGGTAGGCAGG 1167
          ::|||      :::|||::::::
     608 lyLeuIleProAlaProSerMet..... 615

    1166 ATGTCTAGCGGCATTACGCGCTCGTAAGTACCAATCGGCACCATGGCGCG 1117
                                |||||   :::|||   :::::
     616 .........................ValPro...AlaThrAspThrLy 622

    1116 GTCGCCACCGTTGACGGCTGTCGTGAACTTGAAGAGTTTGTTTTTCAGGA 1067
          ::::||||||         :::   :::      :::          :
     622 sAlaProProThrLeuGlnAlaGluThrAlaThrLysProGlnAlaThrS 639

    1066 AATGGCCGAGGGTCGTACGCGTGATGGAGTATTTGTCCGGCTGCGGCGCA 1017
          ::   |||      :::      ::::::::::|||   |||
     639 erAlaProSerProAlaProLysGlnSerPheLeuPheGlyThrGlnAsn 655

    1016 ACCCAGCCGAACAGCTCTTTGCTGCGGCCTTCTTCGATAACGGAAATCTG 967
          |||:::|||:::|||           |||::::::
     656 ThrSerProSerSer.....ProAlaAla..... 663

     966 ATTGTGGTAGCGTCCCAAATAATCGTGCGCGCCTTGTGTAATCGCGCCGT 917

                    |||      |||   ||||||   ::::::   |||:
     664 .............ProAlaAlaSerSerAlaProProMetPheLysProI 676

     916 TCAATACCGAACCGGAAATCACGCGGTTGTCTGCGTCAACCAATTCGCCC 867
          ::   |||   |||      ::::::   ::::::   |||
     676 lePheThrAlaProProLysSerGluLysGluGlyProThr..... 689

     866 GCAGTAATTTGCGATACTTTCGCACCCAAAACGGTACGCAAGAGGCGTGG 817

     689 ..................................................... 689

     816 TTTGTTGACTTGAGAACCACCCAAAGCAATCACGCGCTCGGTGTTCAGAC 767
                       |||      ::::::|||   ::::::
     690 .............ProProGlyProSerValThrAlaThrAla..... 699

     766 GGCCTGTTGCAAACAAACGTCCGATGGCAATTACATCTTGATAATTGATG 717
          |||   :::::::::   |||
     700 ..ProSerSerSerSerLeuPro..... 706

     716 GTCCAAACGGTTTTGTTTGCACCGACCGGCTCAATGAAATGAATGTGCGT 667
          :::   |||         |||||||||
     707 ThrThrThrSerThrThrAlaProThr..... 715

     666 GCCACTCAAACCG......GCCGGATGCGGGCCGCCGAATTCATGTGTTT 623
               ::::::|||      ::::::   |||||||||   |||   |||
     716 ....PheGlnProValPheSerSerMetGlyProProAlaSer...ValP 730

     622 CGATGTTGGCAGCATTTTCAGACGGCACGTCTGCGCCAGCTGCCTTACAC 573
          :::   |||   |||         |||::::::||||||:::
     730 roLeuProAlaProPhePheLysGlnThrThrThrProAlaThrAlaPro 746

     572 ACATGGATTTTACGCTCGGTCAAACGGCTCAATACCAGCANACNTCGTCT 523
          |||
     747 Thr..... 747

     522 GAAATCNTCGNCGGCTTCTTTGATCACAACCACAGGGTCTGCCGCNAGCG 473
                ::::::|||   |||:::|||      :::|||||||:::|||:
     748 .......ThrThrAlaProLeuPheThrGlyLeuAlaSerAlaThrSerA 762

     472 GATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCG... 426
          ::::::::   |||:::   ::::::   ::::::   |||||||||
     762 laValAlaProIleThrSerAlaSerProSerThrAspSerAlaSerLys 778

     425 ...ACGGCAGGGATTTTGCTGAACGGACGGNTACGCAGCGCAGTCCACAA 379
            :::   |||:::   :::|||:::   :::   |||:::|||   ::
     779 ProAlaPheGlyPheGlyIleAsnSerValSerSerSerSerVal...Se 794

     378 ACCGGATTGGATCAGATTGCNNCNNANTTCNTCGCCGCTTAAGTTTGCC. 330
          :                  :::::::::|||:::|||:::   |||:::
     794 rThrThrThrSerThrAlaThrAlaAlaSerGlnProPheLeuPheGlyA 811

     329 .....AACGCTTCGGGCGCGTAGCGTTCGAACTCGATTTCGTCGTTG... 288
               :::||||||:::|||      :::   :::::::::|||:::
     811 laProGlnAlaSerAlaAlaSerPheThrProAlaMetGlySerIlePhe 827

     287 ...........CCTTCAACGGCAATCACGACCGACTGAAGTACGCGCTT 250
                     |||   :::      ||||||         |||   ||
     828 GlnPheGlyLysProProAlaLeuProThrThrThrThrValThrThrPh 844

     249 TTCGCCGCGATGGATGGCGGCGATTTTGCCTGAAACNGGCGCGGTAAACA 200
          ||||           :::|||:::         |||::::::     :
     844 eSerGlnSerLeuHisThrAlaValProThrAlaThrSerSerSerAlaA 861

     199 CCACGCCCGGATNCTTTTTGTCTTCAAACAGCACT...TGGCCTTTTTTG 153
          ::      :::   |||   |||:::   :::|||      |||
     861 laAspPheSerGlyPheGlySerThrLeuAlaThrSerAlaProAlaThr 877

     152 ACGGCATCGCCTTCCTTGACTTTCATCNAGGGGCGCATACCGGCATATTC 103
          :::::::::|||:::|||||||||               |||:::::::::
     878 SerSerGlnProThrLeuThrPheSerAsnThrSerThrProThrPheAs 894

     102 TTCGCCAAGCAACGCGACTTCGGTAATGACGGGCCCGTCATAAATGACTT 53
          :   |||   ::::::::::::            ||||||         :
```

```
    894 nIleProPheGlySerSerAlaLysSerProLeuProSerTyrProGlyA 911

     52 GCTCCGGTCTGCCCGCGATGGGC 30
        :::::      ||||||:::|||
    911 laAsnProGlnProAlaPheGly 918

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU39645

seq_documentation_block:
ID   AAU39645 standard; Protein; 338 AA.
XX
AC   AAU39645;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Propionibacterium acnes immunogenic protein #541.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
XX
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US12865.
XX
PR   21-APR-2000; 2000US-199047P.
PR   02-JUN-2000; 2000US-208841P.
PR   07-JUL-2000; 2000US-216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW,  Persing DH,  Mitcham JL,  Wang SS,  Bhatia A;
PI   L'maisonneuve J,  Zhang Y,  Jen S,  Carter D;
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59508.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris -
XX
PS   Example 1; SEQ ID No 840; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA).
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   338 AA;

alignment_scores:
                Quality:  112.00              Length:    375
                  Ratio:  0.747                 Gaps:     18
  Percent Similarity:  40.000    Percent Identity: 22.667

alignment_block:
 US-09-303-518D-127/rev x AAU39645   ..

 Align seg 1/1  to: AAU39645  from: 1  to: 338

   1031 TCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCCTTCTTC 982
        :::||||||:::::           ||||||     |||   |||
     17 ThrGlyCysAlaGlyLeuThrSerLeuSerSerThrThrArgAlaSerHi 33

    981 GATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCGTGCGCGCCTT 932
                                                |||||||:::|||
     33 sSer..............................SerCysValProV 39

    931 GTGTAATCGCGCCGTTCAATACCGAACCGGAAATCACGCGGTTGTCTGCG 882
                              |||:::|||      :::   |||||||:::
     39 al.................ThrSerProGlySerSerSerLeuSerSer 49

    881 TCAACCAATTCGCCCGCAGTAATTTGCGATACTTTCGCACCCAAAACGGT 832
          |||::::::   |||:::            :::   |||::::::
     50 ProThrThrThrSerAlaIleAlaArgArgArgLeuArgProArgSerPr 66

    831 ACGCAAG.....................AGGCGTGGTTTGTTGACTTGAG 803
           :::                      ||||||:::::   :::
     66 oThrGlnLeuThrCysSerGlyTrpProArgArgSerIleProAlaAlaP 83

    802 AACCACCCAAAGCAATCACG.......................CGCTCG 777
           |||   :::   |||                          :::|||
     83 roAlaProThrSerProThrThrArgAlaArgAsnThrProProGlnSer 99

    776 GTGTTCAGACGGCCTGTTGCAAACAAACGTCCGATGGCAATTACATCTTG 727
        :::   ||||||      ::::::      |||:::      ||||||
    100 AlaThrArgArgPheProThrSerLeuTrpProIleArgThrThrSer.. 115

    726 ATAATTGATGGTCCAAACGGTTTTGTTTGCACCGACCGGCTCAATGAAAT 677
                   :::   ::      |||      |||
    116 .............ArgCysLeuAlaThrAlaArgProGly.......... 124

    676 GAATGTGCGTGCCACTCAAACCGGCCGGATGC......GGGCCGCCGAAT 633
              :::            :::   |||      ||||||   :::
    125 ..TyrSerIleSerThrThrSerSerTrpCysArgValGlyProSerSer 140

    632 TCATGTGTTTCGATGTTGGCA............................ 612
        :::||||||:::   |||:::
    141 AlaCysValThrAlaLeuSerAlaGlyIleThrValThrThrArgSerGl 157

    611 ....GCATTTTCAGACGGCACGTCTGCGCCAGCTGCCTTA...CACACAT 569
           |||::::::   :::||||||   |||  |||           ||||
    157 uCysAlaLeuThrProSerThrSerProProProAlaThrGlyPheThrT 174

    568 GGATTTTACGCTCGGTCAAACGGCTCAATACCAGCANACNTCGTCTGAAA 519
        ||
    174 rp.......................................... 174

    518 TCNTCGNCGGCTTCTTTGATCACAACCACAGGGTCTGCCGCNAGCGGATT 469
              :::::::::::::::|||||:::|||:::
    175 ......SerThrThrMetLeuThrThrSerGlyThr.............. 184

    468 GGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCGACGGCAG 419
           |||   |||||||||   :::   :::          |||:::|||
    185 ....SerArgAlaLeuThrAlaIleArgArgLeuTyrIleSerAlaAlaA 200

    418 GGATTTTGCTGAACGGACGGNTACGC..................... 393
              :::   |||      |||
    200 rgTyrProValValGlySerArgArgPheProArgProTyrArgThrPro 216

    392 ......................AGCGCAGTCCACAAACCGGATTGGATCAG 364
                                    :::   :::|||   |||:::
    217 SerCysGlyAlaMetProThrLeuArgIleSerArgProThrTrpMet.. 232
```

```
 363 ATTGCNNCNNANTTCNTCGCCGCTTAAGTTTGCCAACGCTTCGGGCGCGT 314
                      |||         :::      :::   :::
 233 ..................ProSerMetThrSerCysIleGlyThrThrA 243

 313 AGCGTTCGAACTCGATTTCGTCGTTGCCTTCAACGGCAATCACGACCGAC 264
       |||::::::   |||   ||||||      ||||||   |||
 243 rgArgGlySerProIleProSerLeu......ThrAlaSerThrIleVal 257

 263 TGAAGTACGCGCTTTTCGCCGCGATGGATG............GCGGCGAT 226
      |||:::|||   |||   ||||||:::              ::::::
 258 ArgSerSerArgArgSerValArgTrpValArgSerSerSerSerSerPr 274

 225 TTTGCCTGAAACNGGCGCGGTAAACACCACGCCCGGATNC......TTTT 182
       |||:::|||:::   ::::::||||||   |||:::      |||
 274 oHisProArgThrSerProThrGlyThrThrIleGlySerThrArgPheA 291

 181 TGTCTTCAAACAGCACTTGGCCTTTTTTGACGGCATCGCCTTCCTTGACT 132
       |||      :::|||
 291 rgProSerAlaArgSerTrp.............................. 297

 131 TTCATCNAGGGGCGCATACCGGCATATTCTTCGCCAAGCAACGCGACTTC 82
                                  ||||||||||||:::||||||::
 298 ..........................SerSerProSerSerAlaThrTh 305

  81 G............GTAATGACGGGC 69
     :            ||||||||||||
 305 rSerArgAsnGlyValMetThrGly 313
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY31745

seq_documentation_block:

```
ID   AAY31745 standard; Protein; 430 AA.
XX
AC   AAY31745;
XX
DT   22-NOV-1999  (first entry)
XX
DE   Mycobacterium tuberculosis specific DNA-encoded polypeptide.
XX
KW   Tuberculosis; infection; diagnosis; DNA probe.
XX
OS   Mycobacterium tuberculosis.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 4
FT                   /note= "encoded by TGA"
FT   Misc-difference 6
FT                   /note= "encoded by TGA"
FT   Misc-difference 20
FT                   /note= "encoded by TGA"
FT   Misc-difference 29
FT                   /note= "encoded by TGA"
FT   Misc-difference 54
FT                   /note= "encoded by TGA"
FT   Misc-difference 64
FT                   /note= "encoded by TGA"
FT   Misc-difference 69
FT                   /note= "encoded by TGA"
FT   Misc-difference 89
FT                   /note= "encoded by TGA"
FT   Misc-difference 99
FT                   /note= "encoded by TGA"
FT   Misc-difference 114
FT                   /note= "encoded by TGA"
FT   Misc-difference 119
FT                   /note= "encoded by TGA"
FT   Misc-difference 129
FT                   /note= "encoded by TGA"
FT   Misc-difference 159
FT                   /note= "encoded by TGA"
FT   Misc-difference 169
FT                   /note= "encoded by TGA"
FT   Misc-difference 182
FT                   /note= "encoded by TAG"
FT   Misc-difference 185
FT                   /note= "encoded by TGA"
FT   Misc-difference 219
FT                   /note= "encoded by TGA"
FT   Misc-difference 259
FT                   /note= "encoded by TGA"
FT   Misc-difference 269
FT                   /note= "encoded by TGA"
FT   Misc-difference 291
FT                   /note= "encoded by TGA"
FT   Misc-difference 323
FT                   /note= "encoded by TGA"
FT   Misc-difference 339
FT                   /note= "encoded by TGA"
FT   Misc-difference 349
FT                   /note= "encoded by TGA"
FT   Misc-difference 356
FT                   /note= "encoded by TGA"
FT   Misc-difference 366
FT                   /note= "encoded by TGA"
FT   Misc-difference 410
FT                   /note= "encoded by TGA"
FT   Misc-difference 421
FT                   /note= "encoded by TGA"
XX
PN   EP945462-A1.
XX
PD   29-SEP-1999.
XX
PF   25-MAR-1998;   98EP-0302287.
XX
PR   25-MAR-1998;   98EP-0302287.
XX
PA   (COUL ) CSIR COUNCIL SCI IND RES.
XX
PI   Kumar D,  Srivastava BS,  Srivastava R;
XX
DR   WPI; 1999-530042/45.
DR   N-PSDB; AAX87940.
XX
PT   New nucleic acid molecules, useful for detecting and identifying
PT   Mycobacterium tuberculosis
XX
PS   Disclosure; Page 15-27; 42pp; English.
XX
CC   The present sequence represents an amino acid sequence deduced
CC   from the Mycobacterium tuberculosis specific DNA fragment provided
CC   in AAX87940.  This DNA fragment comprises a StuI-StuI fragment of
CC   M. tuberculosis genomic DNA and contains an insertion sequence-like
CC   element and repetitive sequences.  The DNA fragment is useful as a
CC   probe, especially for detecting or identifying M. tuberculosis in
CC   clinical isolates and body fluids e.g. sputum, cerebrospinal fluid,
CC   pleural fluid, urine, gastric lavage, bronchial lavage, pericardial
CC   or lymph node aspirate (all claimed).  It is also useful for
CC   restriction fragment length polymorphism analysis of M. tuberculosis
CC   isolates (claimed).  The probe provides rapid and specific diagnosis
CC   of tuberculosis and M. tuberculosis infection.
XX
SQ   Sequence   430 AA;
```

alignment_scores:

```
            Quality:  110.00              Length:    446
              Ratio:  0.625                 Gaps:     23
 Percent Similarity:  39.462   Percent Identity: 23.767
```

alignment_block:

US-09-303-518D-127 x AAY31745 ..

Align seg 1/1 to: AAY31745 from: 1 to: 430

```
     44 GACCGGAGCAAGTCATTTATGACGGGCCCGTCATTA.........CCGAA 84
```

```
 34 AspArgGluIleAspGlyGluSerGlyProGlyTyrArgProProProAr 50

 50 gSerHisCys***Ser.ArgProCysArg.............LysArg 61

 62 ArgTyr***ArgCysArgAsp***ArgSerArgCysTrpCysHisProGl 78

 78 yTyrSerCysArgTyrTrpCysHisProArg***SerCysArgCysTrpC 95

 95 ys.TyrArgArg***SerArgArgCysCysSerCysProCysCysArgSe 111

111 rProCys***Ser.....................ArgArgArg***G 120

120 lyCysProCysCysSerCysGlnHis***GlyCysArgTyrCysArgTyr 136

137 ProGlySerArgTyrProSerSerArg.........Cys.........Pr 147

147 oSerLeuArg.........CysArgArgPheArgCysProArg***ArgC 161

161 ysGlnArgTyrTrpCysProAsn***ThrGlyArgCysCysArgCysPro 177

178 SerSerSerArg***GlnTyr...***SerProAlaGlyCysArgArgTh 193

193 rAlaArgCys...ArgCysCysCysCysArgCysTrpArgThrArgCysC 209

209 ysCysCysArgCysTrpGlnSerLeuGly***SerArgProArgSerArg 225

226 SerArgArgCysSerArgArgArgPheGlnAsnArgCysCysArgSerAr 242

242 gGlyPheArgIleArgCysCysSerPheProGlyPhe............A 255

255 rgAsnArgHis***IleLeuArgCysPheHisCysArgTyr...***Ser 270

271 CysArgArgCysArgCysProArgCysPheGlyCys.............. 282

283 ....ArgGlyCysArgCysGlnGlyCys***SerHisArg.......... 294

 34 ..........ArgPheArgCysArgGluCysCysAsnCysArgCysTrp 307

308 ArgCys..................................... 309

310 .ArgGluCysSerArgArgProGlyLeuProGlyArgAsp***ArgProV 326

326 alGlyHisArgLysIleProThrCysCysPheArgCys***ArgSerPro 342

343 ArgSerArgProAlaLeu***TrpProProGlySerCys***ThrAsnPr 359

359 oIleArgCysCysProSer***SerArgProIleProAlaArgProArgL 376

376 euPro......GlyArgSerTyrArgTrpProProThrLysSerGly... 389

390 .........ArgSerGlnAsnCysTrpHisArg................. 397

398 .SerSerGlySerArgThrArgTyrArgHisArg 408
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU35592

seq_documentation_block:

```
ID   AAU35592 standard; Protein; 457 AA.
XX
AC   AAU35592;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Haemophilus influenzae cellular proliferation protein #233.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Haemophilus influenzae.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS53451.
XX
PT   New polynucleotides for the identification and development of
```

```
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 11185; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   457 AA;


alignment_scores:
            Quality:  109.50              Length:    405
              Ratio:  0.545                Gaps:     20
 Percent Similarity:  49.630   Percent Identity: 21.728

alignment_block:
 US-09-303-518D-127 x AAU35592   ..

 Align seg 1/1  to: AAU35592  from: 1  to: 457

     253 CGCGTACTTCAGTCGGTCGTGATTGCCGTTGAAGGCAACGACGAAATCGA 302
         ::::::||||||:::   :::|||          |||:::::::::::::
      11 GlnIleLeuGlnAlaLysLeuIle.........GlyAspGluAsnValGl 24

     303 GTTCGAACGCTACGCGCCCGAAGCGTTGGCAAACTTAAGCGGCGANGAAN 352
         :   |||:::          ::::::       ::::::|||
      24 nValGluLysIleAsnThrAspThrArgLysSerValSer.......... 37

     353 TNNGNNGCAATCTGATCCAATCCGGTTTGTGGACTGCGCTGCGTANCCGT 402
                              ::::::|||:::   ||||||:::   :::
      38 ....................AsnSerLeuPhePheAlaLeuLysGlyGlu 47

     403 CCGTTCAGCAAAATCCCTGCCGTCGAT......GCCGAGCCGTTCGCCAT 446
            |||:::             :::|||      ::::::      |||::
      48 LysPheAspAlaHisGlnTyrLeuAspGlnAlaValSerGlnGlyAlaLe 64

     447 CTTCGTCAATGCGATGGACACCAATCCGCTNGCGGCAGACCCTGTGGTTG 496
         :   :::   :::::::::   |||          :::   |||   :::|
      64 uAlaLeuValValGlnGlnGluAsnSerSerIleSerValProGlnLeuV 81

     497 TGATCAAAGAAGCCGNCGANGATTTCAGACGANGTNTGCTG......... 537
         ||:::|||::::::              |||   :::|||
      81 alValLysAspThr..............ArgIleAlaLeuGlyGluLeu 92

     538 .........GTATTGAGCCGTTTGACCGAGCGTAAAATCCATGTGTGTAA 578
                 :::   ::::::::::::   |||   :::   :::
      93 AlaLysTrpLeuArgGluLysIleAsnProArgThrValAlaMetThrGl 109

     579 GGCAGCTGGCGCAGACGTGCCGTCTGAAAATGCTGCCAACATCGAAACAC 628
          ::::::|||      :::   :::|||   :::|||:::|||      |
     109 ySerSerGlyLysThrThrValLysGluMetThrAlaSerIleLeuGlnH 126

     629 ATGAATTCGGCGGCCCGCATCCGGCCGGTTTGAGTGGCACGCACATTCAT 678
         ||          :::             ::::::      |||:::   :::
     126 isThr......AlaAlaAspSerGluAlaValLeuPheThrAsnGlyAsn 140

     679 TTCATTGAGCCGGTCGGTGCAAACAAAACCGTTTGGACCATCAATTATCA 728
         |||   :::   :::|||:::      |||:::      ::::::   ::
     141 PheAsnAsnAspIleGlyValProLeuThrLeuLeuArgLeuThrGluLy 157

     729 AGATGTAATTGCCATCGGACGTTTGTTTGCAACAGGCCGTCTGAACACCG 778
         :      :::|||
     157 sHisArgPheAla..................................... 161

     779 AGCGCGTGATTGCTTTGGGTGGT.........TCTCAAGTCAACAAACCA 819
               ||||||   ||||||:::         :::::::::|||
     162 .....ValIleGluLeuGlyAlaAsnHisGlnAsnGluIleAsnTyrThr 176

     820 CGCCTCTTGCGTACCGTTTTGGGTGCGAAAGTATCGCAAATTACTGCGGG 869
              |||            :::|||   ::::::::::|||:::   ::
     177 ThrLysLeuValGlnProAsnAlaAlaLeuIleAsnAsnIleAlaProAl 193

     870 CGAATTG.........................GTTGACGCAGACAACC 892
         ::::|||                         |||:::|||   ::::
     193 aHisLeuGluGlyPheGlySerLeuAlaGlyValValGlnAlaLysGlyG 210

     893 GCGTGATTTCCGGTTCGGTATTGAACGGC...GCGATTACACAAGGCGCG 939
         :::::      |||   :::   ||||||   ||||||   ::::::
     210 luIleTyrArgGlyLeuThrLysAsnGlyValAlaIleIleAsnAlaGlu 226

     940 CACGATTATTTGGGACGCTACCACAATCAGATTTCC......GTTATCGA 983
         |||:::::::|||      ::::::::::::||||||      :::|||::
     227 HisAsnHisLeuAspIleTrpGlnLysGluIleSerAsnHisAlaIleGl 243

     984 AGAAGGCCGCAGCAAAGAG............................... 1002
         :      ::::::|||:::
     243 nTyrPheAsnGlyLysAspTyrSerAlaLysAsnIleHisHisThrSerG 260

    1003 ........CTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATC 1044
                    |||      ::::::||||||      :::   :::|||
     260 lnGlySerThrPheThrLeuIleSerProGln...GlyGluIleGluIle 275

    1045 ACGCGTACGACCCTCGGC...CATTTCCTGAAAAACAAACTCTTCAAGTT 1091
         |||          ||||||   |||   :::||||||   |||
     276 ThrLeuProTyrLeuGlyGluHisAsnValLysAsnAlaLeuAlaAlaTh 292

    1092 CACGACAGCCGTCAAC...GGTGGCGACCGCGCCATGGTGCCGATTGGTA 1138
          :::   |||:::|||   |||:::      :::   |||      |||
     292 rAlaLeuAlaMetAsnValGlyAlaThrLeuThrAspValLysAlaGlyL 309

    1139 CTTACGAG.................CGCGTAATGCCGCTAGACATCCTG 1170
               :::                |||:::::::|||::::::::::
     309 euGluGlnArgSerGlnValLysGlyArgLeuProIleGlnValThr 325

    1171 CCTACCCTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCA 1220
         |||:::|||||||||   |||            ||||||   :::|||
     326 ProAsnLeuLeuLeuLeuAsp............AspThrTyrAsnAla.. 337

    1221 AGCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCT 1270
                                  :::::::||| ||||||||||||||
     338 .......................AsnLysAsp.SerLeuCysAlaAla 345

    1271 TCGTCTGCCCGGGCAAATACGAATANGGCCCGC...TGTTGCGTAAGGTG 1317
               ::::::      :::   ::::::   |||      ||||||
     346 IleAspValLeuLysGlyTyrAspAlaPheArgIleLeuCysValGlyAs 362

    1318 CTGGAAACCNTTG 1330
            |||   |||
     362 pMetLysGluLeu 366

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG04814

seq_documentation_block:
ID   ABG04814 standard; Protein; 729 AA.
XX
AC   ABG04814;
XX
DT   13-FEB-2002  (first entry)
XX
```

```
DE   Novel human diagnostic protein #4805.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS69001.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 35173; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   729 AA;


alignment_scores:
           Quality:  108.50            Length:    450
             Ratio:  0.512               Gaps:     19
 Percent Similarity:  47.111   Percent Identity: 21.778

alignment_block:
 US-09-303-518D-127/rev x ABG04814   ..

 Align seg 1/1  to: ABG04814  from: 1  to: 729

    1283 CCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCAATTCCAA 1234
         ||||||              :::|||:::   :::   :::   ::
      51 ProGlyProLysHisGlnProThrValArgGluLeuAlaArgArgCysAr 67

    1233 GCAACCCAATGCTTGC......GCGCTGTCGGTATCGCCGACGATTAAAT 1190
         :   |||:::||||||      |||   ::::::
      67 gLeuProArgAlaCysHisProAlaGlnAsnLeuLeuGluGlnGlnProL 84

    1189 CGCGCAAAAGCAGGGTAGGCAGGATGTCTAGCGGCATTACGCGCTCGTAA 1140
           |||   |||::::::   |||      |||   :::
      84 euArgProSerHisLeuGlnArgHisLeuSerLeuLeuPro..... 97

    1139 GTACCAATCGGCACCATGGCGCGGTCGCCACCGTTGACGGCTGTCGTGAA 1090
           |||:::|||            :::||||||      |||
      98 ...ProLeuGlyLeuSerGlnSerGlyProProGlyLeuAlaProGlnPr 113

    1089 CTTGAAGAGTTTGTTTTTCAGGAAATGGCCGAGGGTCGTACGCGTGATGG 1040
          |||   :::|||                          ::::::   :
     113 oLeuLeu***Leu.........................GlnThrProA 121

    1039 AGTATTTGTCCGGCTGC.............................. 1023
         ::::::::::|||:::
     121 spHisPheAlaGlyAlaAspProCysSerIleHisGlyThrSerHis*** 137

    1022 ......GGCGCAACCCAGCCGAAC......AGCTCTTTGCTGCGGCCTTC 985
               |||:::::::::|||:::      ::::::          |||||
     138 HisGlnGlyThrSerAsnProSerGlyArgAspGlyTyrGlnThrProSe 154

     984 TTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCGTGCGCGC 935
         |   |||                                  |||   ||||
     154 rHisIleCysPro.........................SerProAlaP 162

     934 CT.........TGTGTAATCGCGCCGTTCAATACCGAACCGGAAATCACG 894
         ||           :::::::::      ||||||:::|||:::
     162 roLysGlnSerPheLeuPheGlyThrGlnAsnThrSerProSerSerPro 178

     893 CGGTTGTCTGCGTCAACCAATTCGCCCGCAGTAATTTGCGATACTTTCGC 844
                |||:::::::::::::|||   ::::::          |||::
     179 AlaAlaProAlaAlaSerSerAlaProProMetPheLysProIlePheTh 195

     843 ACCCAAAACGGTACGCAAGAGGCGTGGTTTGTTGACTTGAGAACCACCCA 794
         :                  ::::::::::|||                |||
     195 rAlaProProLysSerGluLysGluGlyProThr......ProProGlyP 210

     793 AAGCAATCACGCGCTCGGTGTTCAGACGGCCTGTTGCAAACAAACGTCCG 744
           ::::::|||   ::::::         |||   :::::::::   |||
     210 roSerValThrAlaThrAla.........ProSerSerSerSerLeuPro 223

     743 ATGGCAATTACATCTTGATAATTGATGGTCCAAACGGTTTTGTTTGCACC 694
                                     :::   |||         |||||
     224 ............................ThrThrThrSerThrThrAlaPr 231

     693 GACCGGCTCAATGAAATGAATGTGCGTGCCACTCAAACCG......GCCG 650
         ||||                          :::::|||      ::::
     231 oThr.........................PheGlnProValPheSerS 239

     649 GATGCGGGCCGCCGAATTCATGTGTTTCGATGTTGGCAGCATTTTCAGAC 600
         ::   |||||||||   |||   |||   :::   |||   |||
     239 erMetGlyProProAlaSer...ValProLeuProAlaProPhePheLys 254

     599 GGCACGTCTGCGCCAGCTGCCTTACACACATGGATTTTACGCTCGGTCAA 550
           |||::::::||||||:::      |||
     255 GlnThrThrThrProAlaThrAlaProThr..................... 264

     549 ACGGCTCAATACCAGCANACNTCGTCTGAAATCNTCGNCGGCTTCTTTGA 500
                                        :::::|||   |||:
     265 ..............................ThrThrAlaProLeuP 270

     499 TCACAACCACAGGGTCTGCCGCNAGCGGATTGGTGTCCATCGCATTGACG 450
         ::|||      :::||||||:::|||:::::::::   |||:::   :::
     270 heThrGlyLeuAlaSerAlaThrSerAlaValAlaProIleThrSerAla 286

     449 AAGATGGCGAACGGCTCGGCATCG......ACGGCAGGGATTTTGCTGAA 406
         :::   ::::::   |||||||||      :::   |||:::   :::||
     287 SerProSerThrAspSerAlaSerLysProAlaPheGlyPheGlyIleAs 303

     405 CGGACGGNTACGCAGCGCAGTCCACAAACCGGATTGGATCAGATTGCNNC 356
         |:::   :::   |||:::|||   :::                   ::::
     303 nSerValSerSerSerSerVal...SerThrThrThrSerThrAlaThrA 319

     355 NNANTTCNTCGCCGCTTAAGTTTGCC......AACGCTTCGGGCGCGTAG 312
```

```
      :::::|||:::|||:::   |||:::      :::|||||||:::|||
  319 laAlaSerGlnProPheLeuPheGlyAlaProGlnAlaSerAlaAlaSer 335

  311 CGTTCGAACTCGATTTCGTCGTTG.............CCTTCAACGGC 277
         :::   ::::::::::|||:::              |||   :::
  336 PheThrProAlaMetGlySerIlePheGlnPheGlyLysProProAlaLe 352

  276 AATCACGACCGACTGAAGTACGCGCTTTTCGCCGCGATGGATGGCGGCGA 227
        ||||||         |||   ||||||            :::|||:
  352 uProThrThrThrThrValThrThrPheSerGlnSerLeuHisThrAlaV 369

  226 TTTTGCCTGAAACNGGCGCGGTAAACACCACGCCCGGATNCTTTTTGTCT 177
      ::        |||::::::     :::      :::   |||  |||
  369 alProThrAlaThrSerSerSerAlaAlaAspPheSerGlyPheGlySer 385

  176 TCAAACAGCACT...TGGCCTTTTTTGACGGCATCGCCTTCCTTGACTTT 130
      :::   :::|||      |||      :::::::::|||:::||||||||
  386 ThrLeuAlaThrSerAlaProAlaThrSerSerGlnProThrLeuThrPh 402

  129 CATCNAGGGGCGCATACCGGCATATTCTTCGCCAAGCAACGCGACTTCGG 80
      |              |||::::::::::   |||   ::::::::::::
  402 eSerAsnThrSerThrProThrPheAsnIleProPheGlySerSerAlaL 419

   79 TAATGACGGGCCCGTCATAAATGACTTGCTCCGGTCTGCCCGCGATGGGC 30
                |||||         ::::::      ||||||:::|||
  419 ysSerProLeuProSerTyrProGlyAlaAsnProGlnProAlaPheGly 435

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB59201

seq_documentation_block:
ID   ABB59201 standard; Protein; 746 AA.
XX
AC   ABB59201;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 4395.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL03304.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 4395; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   746 AA;


alignment_scores:
           Quality:  107.00              Length:    524
             Ratio:  0.476                Gaps:     21
 Percent Similarity:  42.939   Percent Identity: 20.229

alignment_block:
 US-09-303-518D-127/rev x ABB59201   ..

 Align seq 1/1  to: ABB59201  from: 1  to: 746

 1334 TTCTCAANGGTTTCCAGCACCTTACGCAACAGCGGGCCNTATTCGTATTT 1285
      ||||||      |||:::||||||   :::|||:::   ::::::   ||
   53 PheSerHisGluSerThrThrLeuProThrSerAlaLeuPheLys...Le 68

 1284 GCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGTCCAATTCC. 1236
      |      ::::::::::|||   ::::::   ||||||:::|||::::::
   68 uThrThrGluSerSerLeuSerSerSerSerSerSerGlySerSerThrA 85

 1235 ..........................AAGCAACCCAATGCTTGCGCG 1215
                                 |||:::   :::
   85 laHisHisHisThrThrAsnGluLeuLeuLysAsnValThrGlnSerTyr 101

 1214 CTGTCGGTATCGCCGACGATTAAATCGCGCAAAAGCAGGGTAGGCAGGAT 1165
      ||||||::::::              |||   :::      |||   ::
  102 LeuSerThrAlaLysProAlaProValArgThrThrSerAsnGlyThrLe 118

 1164 GTCTAGCGGCATTACGCGCTCGTAAGTACCAATCGGCACCATGGCGCGGT 1115
      :|||:::                                 :::|||
  118 uSerThr.....................................ThrArgA 123

 1114 CGCCACCGTTGACGGCTGTCGTGAACTTGAAGAGTTTGTTTTTCAGGAAA 1065
        |||||||||:::   :::::::::   :::|||   :::
  123 rgProProLeuAlaHisThrThrHisLysAsnSerHisTyr..... 136

 1064 TGGCCGAGGGTCGTACGCGTGATGGAGTATTTGTCCGGCTGCGGCGCAAC 1015
             |||||||||:::           |||:::::::   :::||
  137 .........ValValArgThrThrMetLysProSerAlaAlaProThrTh 150

 1014 CCAGCCGAACAGCTCTTTGCTGCGGCCTTCTTCGATAACGGAAATCTGAT 965
      |      :::   ::::::      |||         |||
  150 rThrIleThrProThrValLysProProArgArgArgThrThrThr.... 165

  964 TGTGGTAGCGTCCCAAATAATCGTGCGCGCCTTGTGTA..... 927
             :::|||      :::   ||||||   |||
  166 .....AlaLysProValLeuThrThrAlaProIleValGluGlnArgIle 180

  927 ..... 927

  181 GluThrThrSerValProThrLysPheValThrPheGlnIleValGluTh 197

  926 .ATCGCGCCGTTCAATACCGAACCG..... 903
         :::|||      |||||||||
  197 rThrThrProAlaValThrGluProThrHisGlnGlnGluThrThrThrT 214

  902 .....GAA 900
                                                    :::
  214 hrThrThrArgArgHisThrArgProThrGlySerSerSerValAlaSer 230

  899 ATCACGCGGTTGTCTGCGTCAACCAATTCGCCCGCAGTAATTTGCGATAC 850
         |||      :::::::::|||::::::   ::::::         ||
  231 ArgThrThrSerLysThrThrThrThrThrThrThrThrThrThrTh 247

  849 TTTCGCACCCAAAACGGTACGCAAGAGGCGTGGTTTGTTGACTTGAGAA. 801
      |   ||||||   |||:::   :::   |||:::      |||   :::
  247 rAlaAlaProProThrThrThrSerThrArgAlaAlaThrThrThrGlnA 264
```

```
 800 ........CCACCCAAAGCAATCACGCGCTCG..................  777
                ||||||          |||   |||
 264 rgProThrProProLeuArgThrThrThrSerLysProHisLysThrHis  280

 776 GTGTTCAGACGGCCTGTTGCAAACAAACGTCCGATGGCAATTACATCTTG  727
           |||||||||:::      |||:::|||::::::   |||
 281 SerSerArgArgProAlaProAlaLysLysProValSerThrThr.....  295

 726 ATAATTGATGGTCCAAACGGTTTTGTTTGCACCGACCGGCTCAATGAAAT  677
                     |||:::      :::   |||   |||   :::
 296 ................ThrThrThrAlaSerGlyThrGlnSerThrArg.  306

 676 GAATGTGCGTGCCACTCAAACCGGCCGGATGCGGGCCGCCGAATTCATGT  627
                      |||||||||                   :::
 307 ................LysProAla.................AspSer  311

 626 GTTTCGATGTTGGCAGCATTTTCAGACGGCACGTCTGCG.........CC  586
     ::::::      ::::::   ::::::||||||:::|||          ||
 312 LeuAlaSerGlySerThrThrAlaAsnGlyThrThrAlaThrThrLeuPr  328

 585 AGCTGCCTTACACACATGGATTTTACGCTCGGTCAAACGGCTCAATACCA  536
     ||||                     |||   |||   :::   ::::::|
 328 oAlaArg.................ArgProValThrGlnThrSerSerS  339

 535 GCANACNTCGTCTGAAATCNTCGNCGGCTTCTTTGATCACAACC......  492
     ||::::::         |||:::::::::::::::   ||||||
 339 erSerThrThrThrThrSerThrThrSerThrIleThrThrThrValPhe  355

 492 ..............................................:.  492

 356 LysAspGluLeuProLysAsnArgThrThrThrArgLeuProGluArgTh  372

 491 ............................ACAGGG.........TCTG  482
                                  ||||||         ::::
 372 rThrGlnAlaProArgProArgProLysProThrGlyGluIleValLysV  389

 481 CCGCNAGCGGATTGGTGTCCATCGCATTGACGAAGATGGCGAACGGCTCG  432
     ::   :::   |||::::::   :::   |||   :::::::::||||||
 389 alProAlaLeuLeuAlaGluProThrThrThrThrValThrLysGlySer  405

 431 GCATCGACGGCAGGGATTTTGCTGAACGGACGGNTACGCAGCGCAGTCCA  382
     :::|||::::::                        |||:::
 406 SerSerSerSer.........................SerSerSerPr  413

 381 CAAACCGGATTGGATCAGATTGCNNCNNANTTCNTCGCCGCTT.......  339
      ||||||              ::::::::::::::::||||||
 413 oLysProPro............AlaSerSerThrThrProLeuValThrL  426

 338 ..AAGTTTGCCAACGCTTCGGGCGCGTAGCGTTCGAACTCGATTTCGTCG  291
       |||      ::::::::::::|||      :::::::::   ::::::
 426 ysLysLysProThrThrThrAsnAlaProThrThrThrThrThrThrThr  442

 290 TTGCCTTCAACGGCAATCACGACCGACTGAAGTACGCGCTTTTCGCCGCG  241
           :::|||:::   ||||||        ||||||   :::
 443 ThrThrThrThrThrThrThrProLysProThrArgArgThrLysPr  459

 240 ATGGATGGCGGCGATTTTGCCTGAAACNGGCGCGGTAAACACCACGCCCG  191
                      |||   |||   :::::::::||||||   :
 459 o.................ProThrThrThrThrThrThrThrLysA  470

 190 GATNCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGACGGCATCGCCT  141
     :::::      ::::::::::::|||           |||||||:::|||
 470 laThrThrThrThrThrLysAlaThrThrThrThrThrAlaLysPro  486

 140 TCCTTGACTTTCATCNAGGGGCGCATACCGGCATATTCTTCGCCAAGCAA  91
        :::|||                  |||      |||:::|||   ::
 487 LeuIleThr............ThrGluProProThrSerAlaProLeuTh  499

  90 CGCGACTTCGGTAATGACGGGC  69
     ::::|||:::      ||||||
 499 rThrThrThrLysLysThrGly  506
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB64198

```
seq_documentation_block:
ID   ABB64198 standard; Protein; 2406 AA.
XX
AC   ABB64198;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 19386.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL08301.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 19386; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2406 AA;


alignment_scores:
           Quality:  107.00            Length:    505
             Ratio:  0.502               Gaps:     25
 Percent Similarity:  42.178   Percent Identity: 20.594

alignment_block:
 US-09-303-518D-127/rev x ABB64198    ..

 Align seg 1/1  to: ABB64198  from: 1  to: 2406

    1298 CCNTAT.................TCGTATTTGCCCGGGCAGACGAAGCT 1267
         ||||||                 ||||||||||||           ::
    1266 ProTyrSerGlyLeuThrHisGlySerTyrLeuProProValLeuProVa 1282

    1266 G.....................................CACAAAGCGAGGT 1253
         :                                     |||::::::   |
    1282 lAlaThrProAsnLeuSerAsnLeuProProThrGlnHisArgSer...S 1298

    1252 CTTCTTCGTCCAATTCCAAGCAA.....................CCC 1227
         ||:::|||   ||||||::::::                     |||
    1298 erAspSerArgAsnSerArgGluSerProAlaSerLeuLysSerThrPro 1314
```

```
1226 AATGCTTGCGCGCTGTCGGTATCG......CCGACGATTAAATCG..... 1188
1315 SerAsnIleGlyLeuAsnValSerMetAlaProThrLeuArgSerIleTh 1331

1187 .CGCAAAAGCAGGGTAGGCAGGATGTCTAGCGGCATTACGCGCTCGTAAG 1139
1331 rProLeuAsnAsnSerSerAlaIleSerSerGlyAlaSerGlnProValV 1348

1138 TACCAATCGGCACCATGGCGCGGTCGCCACCGTTGACGGCTGTCGTGAAC 1089
1348 alSerValValProSerAlaAsnSerThrAlaLeuSer..... 1360

1088 TTGAAGAGTTTGTTTTTCAGGAAATGGCCGAGGGTCGTA.....CG 1048
1361 .................MetSerAsnProHisIleSerHisSerHisHi 1371

1047 CGTGATGGAGTATTTGTCCGGC.....TGCGGCG 1019
1371 sValProAlaTyrAlaSerGlyAlaPheSerSerSerAlaAlaAlaGlyT 1388

1018 CAACCCAGCCGAACAGCTCTTTG......CTGCGGCCTTCTTCGATAACG 975
1388 hrSerThrProAsnSerGlyLeuSerThrLeuAlaValThrSerLeuSer 1404

974 GAAATCTGATTGTGGTAGCGTCCCAAATAATCGTGCGCGCCTTGTGTAAT 925
1405 Thr.....SerAlaAlaPro..... 1409

924 CGCGCCGTTCAATACCGAACCGGAAATCACGCGGTTG..... 888
1410 .GlnProHisSerHisPheProGlnSerThrGlnMetLeuProGlnSerG 1426

887 .....TCTGCGTCA 879
1426 lyAsnPheSerSerValSerHisLeuLeuThrThrHisProMetSerSer 1442

878 ACCAATTCGCCCGCAGTAATTTGCGATACTTTCGCACCCAAAACGGTACG 829
1443 GlnAsnGlnProMetValArgCysGlySer.....ThrLeuTy 1455

828 CAAGAGGCGTGGTTTGTTGACTTGAGAACCACCCAAAGCAATCACGCGCT 779
1455 rSerGlnSerSerAlaAlaAlaThrAlaProProSerAlaAlaAla.... 1470

778 CGGTGTTCAGACGGCCTGTTGCAAACAAACGTCCGATGGCAATTACATCT 729
1471 .............AlaValSerAsnPheThrProServal..... 1479

728 TGATAATTGATGGTCCAAACGGTTTTGTTTGCACCGACCGGCTCAATGAA 679
1480 ......LeuAlaValGlnSerLeuThrThrAlaValThrSerSer..... 1492

678 ATGAATGTGCGTGCCACTCAAACCGGCCGGATGCGGGCCGCCGAATTCAT 629
1492 ..... 1492

628 GTGTTTCGATGTTGGCAGCATTTTCAGACGGCACGTCTGCGCCAGCTGCC 579
1493 .....SerSerSerProSerThr 1498

578 TTACACACATGGATTTTACGCTCGGTC.....AAACG 547
1499 LeuSerSerSerValIleGlnLysValIleSerProLysGlyGluSerPr 1515

546 GCTCAATACCAGCANACNTCGTCTGAAATCNTCGNCGGCTTCTTTGATCA 497
1515 oCysAsnLysAspArgAspSerSerTyrSerSerProAlaAsnAlaValV 1532

496 CAACCACA.....GGGTCTGCC... 480
1532 alThrThrCysAlaProThrThrProIleValSerSerGlySerAlaArg 1548

479 .....GCNAGCGGATTGGTGTC 463
1549 ProThrProProLeuSerAsnCysThrSerMetGlyIleGlyMetValAs 1565

462 CATCGCATTGACGAAGATGGCGAACGGCTCGGCA..... 429
1565 nAlaAlaSerThrAlaArgSerSerCysAsnAlaIleSerProLeuSerI 1582

428 .....TCGACGGCAGGGATTTTGCTGAACGGACGGNTACGCAGCGCAGTC 384
1582 leProAlaThrAlaGlyIleHisValSerAlaThr..... 1593

383 CACAAACCGGATTGGATCAGATTGCNNCNNANTTCNTCGCCGCTTAAGTT 334
1594 ...AsnProSerPheGlnSerSerSerTyrPheProThrProLeuAlaPr 1609

333 TGCCAACGCTTCGGGCGCGTAGCGTTCGAACTCGATTTCGTCGTTGCCTT 284
1609 oProProSerSerProSerProAlaThrSerSerAlaAlaIleIleSerS 1626

283 CAACGGCAATCACGACCGACTGAAGTACGCGCTTTTCGCCG..... 243
1626 erSerAla.....SerGlnPheAsnProAlaValSer 1636

242 CGATGGATGGCGGCGATTTTGCCTGAAACNGGCGCGGTAAACACCACGCC 193
1637 HisSerMetSerSerIleValThrThrAlaGlyAlaThrThrThrThrAl 1653

192 CGGATNCTTTTTGTCTTCAAACAGCACTTGGCCTTTTTTGACGGCATCGC 143
1653 aSerSer.....ValThrGlnP 1659

142 CTTCCTTGACTTTCATCNAGGGGCGCATACCGGCATATTCTTCGCCAAGC 93
1659 roSerValAlaAlaIleSerAsnProValThrAsnThrProHisProPhe 1675

92 AACGCGACTTCGGTA 78
1676 SerAlaGluSerLeu 1680
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23390

seq_documentation_block:

ID ABG23390 standard; Protein; 1209 AA.
XX
AC ABG23390;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #23381.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS87577.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in

```
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 53749; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1209 AA;


alignment_scores:
           Quality:  106.50              Length:    452
             Ratio:  0.552                Gaps:     17
 Percent Similarity:  42.699   Percent Identity: 21.460

alignment_block:
 US-09-303-518D-127/rev x ABG23390   ..

 Align seg 1/1  to: ABG23390  from: 1  to: 1209

   1313 TTACGCAACAGCGGGCCNTATTCGTATTTGCCCGGGCAGACGAAGCTGCA 1264
        |||   :::||||||||||   :::   ||||||
    596 LeuSerGlnSerGlyProProGlyLeuLeuPro..................  606

   1263 CAAAGCGAGGTCTTCTTCGTCCAATTCCAAGCAACCCAATGCT...TGCG 1217
               |||   |||   :::||||||   |||:::::      :
    607 ..........SerProSerPheAspSerLysProProThrThrLeuLeuG  620

   1216 CGCTGTCGGTATCGCCGACGATTAAATCGCGCAAAAGCAGGGTAGGCAGG 1167
        ::|||      :::|||::::::
    620 lyLeuIleProAlaProSerMet............................  627

   1166 ATGTCTAGCGGCATTACGCGCTCGTAAGTACCAATCGGCACCATGGCGCG 1117
                              ||||||   :::|||   :::::
    628 ..........................ValPro...AlaThrAspThrLy  634

   1116 GTCGCCACCGTTGACGGCTGTCGTGAACTTGAAGAGTTTGTTTTTCAGGA 1067
        ::::||||||         :::   :::      :::            :
    634 sAlaProProThrLeuGlnAlaGluThrAlaThrLysProGlnAlaThrS  651

   1066 AATGGCCGAGGGTCGTACGCGTGATGGAGTATTTGTCCGGCTGCGGCGCA 1017
        ::   |||      :::      :::::::::|||   |||         .
    651 erAlaProSerProAlaProLysGlnSerPheLeuPheGlyThrGlnAsn  667

   1016 ACCCAGCCGAACAGCTCTTTGCTGCGGCCTTCTTCGATAACGGAAATCTG 967
        |||:::|||:::|||           |||::::::
    668 ThrSerProSerSer............ProAlaAla..............  675

    966 ATTGTGGTAGCGTCCCAAATAATCGTGCGCG.............CCTT 932
                   |||      |||   |||              |||
    676 .............ProAlaAlaSerSerAlaSerProMetPheLysProI  688

    931 GTGTAATCGCGCCGTTCAATACCGAA..................... 906
              ||||||   ::::::|||
    688 lePheThrAlaProProLysSerGluLysGluGlyProThrProProGly  704

    905 CCGGAAATCACGCGGTTGTCTGCGTCAACCAATTCGCCCGCAGTAATTTG 856
        |||::::::|||      :::   |||::::::|||      :::
    705 ProSerValThrAlaThrAlaProSerSerSerSerLeuProThrThrTh  721

    855 CGATACTTTCGCACCCAAAACGGTACGCAAGAGGCGTGGTTTGTTGACTT 806
         :::|||   ||||||
    721 rSerThrThrAlaPro.....................................  726

    805 GAGAACCACCCAAAGCAATCACGCGCTCGGTGTTCAGACGGCCTGTTGCA 756

    726 .....................................................  726

    755 AACAAACGTCCGATGGCAATTACATCTTGATAATTGATGGTCCAAACGGT 706

    726 .....................................................  726

    705 TTTGTTTGCACCGACCGGCTCAATGAAATGAATGTGCGTGCCACTCAAAC 656
           |||   |||:::   |||
    727 .ThrPheGlnProValPheSer.....................................  733

    655 CGGCCGGATGCGGGCCGCCGAATTCATGTGTTTCGATGTTGGCAGCATTT 606
           :::   |||||||||   |||   |||   :::   |||   |||
    734 .....SerMetGlyProProAlaSer...ValProLeuProAlaProPhe  747

    605 TCAGACGGCACGTCTGCGCCAGCTGCCTTACACACATGGATTTTACGCTC 556
                |||::::::||||||:::      |||
    748 PheLysGlnThrThrThrProAlaThrAlaProThr..............  759

    555 GGTCAAACGGCTCAATACCAGCANACNTCGTCTGAAATCNTCGNCGGCTT 506
                                                  ::::::|||
    760 .....................................ThrThrAlaP  763

    505 CTTTGATCACAACCACAGGGTCTGCCGCNAGCGGATTGGTGTCCATCGCA 456
           |||:::|||      :::|||||||:::|||:::::::::   |||:::
    763 roLeuPheThrGlyLeuAlaSerAlaThrSerAlaValAlaProIleThr  779

    455 TTGACGAAGATGGCGAACGGCTCGGCATCG......ACGGCAGGGATTTT 412
           ::::::   ::::::   |||||||||      :::   |||:::
    780 SerAlaSerProSerThrAspSerAlaSerLysProAlaPheGlyPheGl  796

    411 GCTGAACGGACGGNTACGCAGCGCAGTCCACAAACCGGATTGGATCAGAT 362
         :::|||:::   :::   |||:::|||   :::
    796 yIleAsnSerValSerSerSerSerVal...SerThrThrThrSerThrA  812

    361 TGCNNCNNANTTCNTCGCCGCTTAAGTTTGCC......AACGCTTCGGGC 318
           :::::::::|||:::|||:::   |||:::      :::|||||||:::
    812 laThrAlaAlaSerGlnProPheLeuPheGlyAlaProGlnAlaSerAla  828

    317 GCGTAGCGTTCGAACTCGATTTCGTCGTTG..............CCTTC 283
        |||      :::   :::::::::|||:::                |||
    829 AlaSerPheThrProAlaMetGlySerIlePheGlnPheGlyLysProPr  845

    282 AACGGCAATCACGACCGACTGAAGTACGCGCTTTTCGCCGCGATGGATGG 233
         :::      ||||||      |||   ||||||      :
    845 oAlaLeuProThrThrThrThrValThrThrPheSerGlnSerLeuProT  862

    232 CGGCGATTTTGCCTGAAACNGGCGCGGTAAACACCACGCCCGGATNCTTT 183
        ::|||:::         |||::::::      :::      :::   |||
    862 hrAlaValProThrAlaThrSerSerSerAlaAlaAspPheSerGlyPhe  878

    182 TTGTCTTCAAACAGCACT...TGGCCTTTTTTGACGGCATCGCCTTCCTT 136
           |||:::   :::|||      |||      ::::::::::|||:::||
    879 GlySerThrLeuAlaThrSerAlaProAlaThrSerSerGlnProThrLe  895

    135 GACTTTCATCNAGGGGCGCATACCGGCATATTCTTCGCCAAGCAACGCGA 86
        |||||||                |||::::::::::   |||   :::::::
    895 uThrPheSerAsnThrSerThrProThrPheAsnIleProPheGlySerS  912

     85 CTTCGGTAATGACGGGCCCGTCATAAATGACTTGCTCCGGTCTGCCCGCG 36
```

```
      :::::           ||||||        ::::::      ||||||
  912 erAlaLysSerProLeuProSerTyrProGlyAlaAsnProGlnProAla 928

   35 ATGGGC 30
      :::|||
  929 PheGly 930
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG90820

seq_documentation_block:

```
ID   AAG90820 standard; Protein; 1209 AA.
XX
AC   AAG90820;
XX
DT   26-SEP-2001  (first entry)
XX
DE   C glutamicum protein fragment SEQ ID NO: 4574.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
XX
PF   18-DEC-2000; 2000EP-0127688.
XX
PR   16-DEC-1999;   99JP-0377484.
PR   07-APR-2000; 2000JP-0159162.
PR   03-AUG-2000; 2000JP-0280988.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI   Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR   WPI; 2001-376931/40.
DR   N-PSDB; AAH66039.
XX
PT   Novel polynucleotides derived from Coryneform bacteria, for identifying
PT   mutation point of a gene, measuring expression of a gene, analysing
PT   expression profile or pattern of a gene and identifying homologous gene
PT   -
XX
PS   Claim 17; SEQ ID NO: 4574; 246pp + Sequence Listing; English.
XX
CC   The present invention provides a number of nucleotide and protein
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   are useful for identifying the mutation point of a gene derived from a
CC   mutant of coryneform bacterium, measuring expression amount and
CC   analysing the expression profile or expression pattern of a gene derived
CC   from Coryneform bacterium, and identifying a homologue of a gene derived
CC   from coryneform bacterium. Coryneform bacteria are useful for producing
CC   amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC   particularly L-lysine. The present sequence is a protein described
CC   in the exemplification of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office.
XX
SQ   Sequence   1209 AA;
```

```
alignment_scores:
            Quality:  106.50            Length:    550
              Ratio:  0.431               Gaps:     27
 Percent Similarity:  44.909   Percent Identity: 21.273

alignment_block:
 US-09-303-518D-127 x AAG90820   ..
```

```
Align seg 1/1  to: AAG90820  from: 1  to: 1209

     49 GAGCAAGTCATTTATGACGGGCCCGTCATTACCGAAGTCGCGTTGCTTGG 98
        ::::::|||      |||      :::|||  :::|||:::     ||
    649 LysSerValProMetAspArgValIleIleGlyAspValGly...TyrGl 664

     99 CGAAGAATATGCCGGTATGCGCCCCTNGATGAAA...GTCAAGGAAGGCG 145
        |:::      ::::::::::|||  ::::::|||  |||:::::|||
    664 yLysThrGluValAlaValArgAlaAlaPheLysAlaValGlnAspGlyL 681

    146 ATGCCGTC.........................AAAAAAGGCCAA 165
            |||                          ::::::  |||
    681 ysGlnValAlaValLeuValProThrThrLeuLeuAlaGlnGlnHisGln 697

    166 GTGCTGTTTGAAGACAAAAAGNAT......CCGGGCGTGGTG...... 201
            ||||||::::::  :::      |||  ::::::
    698 SerThrPheGluGluArgMetThrGlyPheProValThrIleLysGlyLe 714

    202 .......TTTACCGCGCCN.................GTTTCAGGCAAAA 226
              ||||||:::|||                 :::||||||
    714 uSerArgPheThrSerProAlaGluSerArgGluIleLeuSerGlyLeuA 731

    227 TCGCCGCC.........ATCCATCGCGGCGAAAAGCGCGTACTTCAGTCG 267
          |||:::         |||      |||      |||:::||||||:::
    731 laAlaGlySerValAspIleValIleGlyThrHisArgLeuLeuGlnThr 747

    268 GTCGTG.......................................... 273
          |||
    748 GlyValGlnTrpLysAsnLeuGlyLeuValIleValAspGluGluGlnAr 764

    274 .ATTGCCGTTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGCGCCCG 322
         ::::::||||||  ::::::::::|||:::     |||        :
    764 gPheGlyValGluHisLysGluHisIleLysAlaLeuArgThrHisValA 781

    323 AAGCGTTGGCAAACTTAAGCGGCGANGAANTNNGNNGCAATCTGATCCAA 372
        :::::|||  ::::::|||:::::::            :::|||  :::
    781 spValLeu...ThrMetSerAlaThrProIleProArgThrLeuGluMet 796

    373 TCCGGTTTGTGGACTGCGCTGCGTANCCGTCCGTTCAGCAAAATCCCTGC 422
        |||         :::::::::|||                      |||
    797 Ser......MetAlaGlyIleArgGluMetThrThrMetLeuThrProPr 811

    423 CGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATC 472
           |||  :::|||:::     :::|||:::     :::  :::
    811 oGluAspArgHisProIleLeuThrTyrValGlyProTyrGluAspLysG 828

    473 CGCTNGCGGCA..................................... 483
          :::||||||
    828 lnValAlaAlaSerIleArgArgGluLeuLeuArgAspGlyGlnValPhe 844

    484 .........GACCCTGTGGTTGTGATCAAAGAAGCCGNCGANGATTTCAG 524
                :::  |||:::  |||::::::  :::  ::::::||
    845 PheIleHisAsnLysValAlaAspIleGluLysLysAlaArgGluIleAr 861

    525 A.................CGANGTNTGCTGGTATTGAGCCGTTTGACCG 556
        |                |||     ::::::  ::::::::::::|
    861 gAspLeuValProGluAlaArgValValValAlaHisGlyGlnMetSerG 878

    557 AGCGTAAAATCCAT................................. 570
        ||:::  ::::::
    878 luGluLeuLeuGluGlnThrValGlnGlyPheTrpAspArgGluTyrAsp 894

    571 ......GTGTGTAAGGCA.........GCTGGCGCAGACGTGCCGTCTGA 605
              ||||||  :::         :::|||  |||:::  :::
    895 ValLeuValCysThrThrIleValGluThrGlyLeuAspIleSerAsnAl 911

    606 AAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCGGCCG 655
         |||:::     :::|||            :::  |||     |
    911 aAsnThrLeuIleValGlu..............AsnAlaHisHisMetG 923

    656 GTTTGAGTGGCACGCAC............................... 672
        ||||||||      |||
    923 lyLeuSerGlnLeuHisGlnLeuArgGlyArgValGlyArgSerArgGlu 939
```

```
 673 .........ATTCATTTCATTGAGCCGGTCGGTGCAAACAAAACCGTTTG 713
                  :::|||:::   |||   ||||||:::   |||
 940 ArgGlyTyrAlaTyrPheLeuTyrProLysGlyAlaThrLeuThrGluMe 956

 714 GACCATCAATTATCAAGATGTAATTGCCATCGGACGTTTGTTTGCAACAG 763
      :::  :::         :::||||||   ::::::   :::::::::|
 956 tSerTyrAspArgLeuAlaThrIleAlaGlnAsnAsnAspLeuGlyAlaG 973

 764 GCCGTCTGAACACCGAGCGCGTGATTGCTTTGGGTGGTTCTCAAGTCAAC 813
     ||       :::   :::   :::   :::   |||:::
 973 lyMetAlaValAlaMetLysAspLeuGluMetArgGlyAla......... 986

 814 AAACCACGCCTCTTGCGTACCGTTTTGGGTGCGAAAGTATCGCAAATTAC 863
                     :::|||||||||||||:::   |||
 987 ..............GlyAsnValLeuGlyAlaGluGlnSerGlyHisIl 998

 864 TGCG..............................GGCGAATTGGTTG 880
      |||                              ||||||   |||:
 998 eAlaGlyValGlyPheAspLeuTyrValArgLeuValGlyGluAlaValG 1015

 881 ACGCAGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGCGCGATT.... 927
     ::|||   :::   :::   :::|||:::|||:::::::|||:::::
1015 luAlaTyrArgAlaLeuAlaAspGlyLysValValAspGlyThrValLys 1031

 928 ........................ACACAAGGCGCGCAC......... 942
                              ||||||
1032 GlyProLysGluIleArgValAspLeuProValAspAlaHisIleProGl 1048

 943 .GATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCC 991
         |||::: :::|||      :::|||:::   :::   :::|||
1048 uLysTyrIle.AsnAlaGluArgLeuArgLeuGluIleTyrArgLysLeu 1064

 992 GCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCC 1041
     |||   ::::::   ::::::   ||||||   |||   :::   ::::
1065 AlaGlnSerGluSerGluValAspLeuArg.LeuAlaValGluGluMetG 1081

1042 ATCACGCGTACGACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTT 1091
     ::      |||   |||   |||          ::::::   |||   |||
1081 luAspArgTyrGlyProIleProGluGluValGluArgLeuLeuAlaVal 1097

1092 CACGACAGCCGTCAACGGTGGCGACCGCGCCAT............GGTGC 1129
           |||:::      |||      |||           :::::
1098 SerArgLeuArgHisLeuMetArgGluAlaHisLeuThrAspIleAlaVa 1114

1130 CGATTGGTACTTACGAGCGCGTAATGCCGCTAGACATCCTGCCTACC... 1176
     :::: ||||||      :::|||   |||:::|||:::      :::
1114 lGln.GlyThrArgIleLysValHisProValAspLeuAlaAspSerGln 1130

1177 ...CTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAAGC 1223
        :::   |||:::   |||:::   |||   |||      |||   ||
1131 GlnValArgLeuLysArgLeuPheProGlyAlaThrTyrArgAla...Al 1146

1224 ATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCGCTTTGTGCAGCTTCG 1273
     |      ::::::::::|||                          ||||||
1146 aAlaLysAlaIleGlnLeu.......................SerPheP 1155

1274 TCTGCCCGGGCAAATACGAATANGGCCCGCTGTTGCGTAAGGTG 1317
               |||:::      :::   ||||||||||||   |||
1155 roLysThrGlyAsnLysValThrAspProLeuLeuArgAspVal 1169
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY04955

seq_documentation_block:

ID AAY04955 standard; Protein; 573 AA.
XX
AC AAY04955;
XX
DT 06-JUL-1999 (first entry)
XX
DE Mycobacterium species protein sequence 41T#3.
XX
KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW hybridisation; detection; vaccine; immunisation; infection.
XX
OS Mycobacterium sp.
XX
PN WO9909186-A2.
XX
PD 25-FEB-1999.
XX
PF 14-AUG-1998; 98WO-FR01813.
XX
PR 11-SEP-1997; 97FR-0011325.
PR 14-AUG-1997; 97FR-0010404.
XX
PA (INSP) INST PASTEUR.
XX
PI Gicquel B, Lim EM, Pelicic V, Portnoi D, Goguet de la Salmoniere Y;
PI Guigueno A;
XX
DR WPI; 1999-181045/15.
DR N-PSDB; AAX34206.
XX
PT Mycobacterial DNA vectors containing reporter constructs - for
PT identifying coding or promoter sequences involved in
PT infection-associated protein expression
XX
PS Claim 32; Fig 41T; 309pp; French.
XX
CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
CC proteins from various Mycobacterium species microorganisms. The
CC encoding nucleotide sequences can be used as primers and probes for
CC methods for detecting and identifying mycobacteria, especially belonging
CC to the M. tuberculosis complex. The encoded proteins can be used in
CC vaccines for immunisation against a bacterial or viral infection.
XX
SQ Sequence 573 AA;

alignment_scores:

Quality:	105.50	Length:	521
Ratio:	0.493	Gaps:	29
Percent Similarity:	41.075	Percent Identity:	24.568

alignment_block:
US-09-303-518D-127 x AAY04955 ..

Align seg 1/1 to: AAY04955 from: 1 to: 573

```
  35 TCGCGGGCAGACCGGAGCAAGTCATTTATGACGGGCCCGTCATTA..... 79
     |||   :::   ||||||   :::   |||   :::   ||||||
  45 SerThrSerThrArgSerThrGlySerMetCysSerArgSerLeuThrPr 61

  80 ....CCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGGTATGCGCCCCTN 125
         |||   |||   |||   :::                |||:::|||
  61 oIleProAlaSerThrCysCysSerPro............CysValPro. 73

 126 GATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTG 175
          :::|||   |||   :::   ::::::      :::
  74 .....ArgSer***LysLeuIleCysThrArgIleArgArgLeuThrPro 88

 176 AAGACA...AAAAGNATCCGGGCGTGGTGTTTACCG............CG 210
     ||||||   :::   :::|||           |||            ||
  89 LysThrTrpArgGlyLeuArgProLeuSerArgProAlaArgValGluAr 105

 211 CCNGTTTCAGGCAAAATCGCCGCCATCCATCGCGGCGAAAAGCGCGTACT 260
     | :::   ||||||   ||||||   |||            ||||||
 105 g.SerSerArgGlnLeuArgArgAspProValProAspProAlaArgAsp 121

 261 TCAGTCGGT...CGTGATTGC.........................CG 280
           |||   |||||||||
 122 ArgArgGlyAspArgAspCysGlyArgArgLeuTrpArgGlyIleAlaGl 138

 281 TTGAAGGCAACGACGAAATCGAGTTCGAACGCTACGCGCCCGAAGCGTTG 330
         |||:::::::|||   |||   |||::::::      |||   ||||
```

```
 138 yGlyArgArgGlnArgProArgGlyArgValValAlaArgArgArgValG 155

 331 GCAAACTTAAGCGGCGANGAANTNNGNNGCAATCTGATCCAATCCGGTTT 380
     ||         ||||||          ::::::          ::::::
 155 ly.........ArgArgValAspGlyTrpArgGlyGlyAlaGlyAsnTyr 168

 381 GTGGACTGCGCTGCGTANCCGTCCGTTCAGCAAAATCCCTGCCGT..... 425
               ||||||          :::::::::|||   |||
 169 ArgLeuGlyAlaAlaGlyArgArgArgSerArgArgProValArgAlaAr 185

 426 ..........CGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGAC 465
               ||||||         |||:::   |||         ||||
 185 gGlyValGlyArgCysGlyHisArgArgArg***Arg......GlyGlyH 200

 466 ACCAA.....................TCCGCTNGCGGCAGACCCTGTGGT 494
     ||:::                   ::::::   :::|||   ||||||
 200 isArgAlaGlyLysAspProArgThrAla***ArgAlaArgArgCysGly 216

 495 TGTGATCAAAGAAGCCGNCGANGATTTCAGACGANGTNTGCTGGTATTGA 544
               |||   ||||||:::      :::   :::||||||
 217 ArgGlyGlyArg...ArgArgThrGlyProAlaGlySerAlaGly.ArgV 232

 545 GCCGTTTGACCGAGCGTAAAATCCATGTGTGTAAGGCAGCTGGCGCAGAC 594
          |||   :::   :::      :::|||   ::::::
 232 alAlaLeuHisHisLeuArgAlaGlyThrCysProGlyGlyLeuArgThr 248

 595 GTGCCGTCTGAAAATGCTGCCAACATCGAAACACATGAATTCGGCGGCCC 644
     :::   ::::::   |||         :::   |||   :::      ||
 249 LeuAlaAlaArgValAla.........AspArgHisGlyTyrProThrPr 262

 645 GCATCCGGCC..............GGTTTGAGTGGCACGCACATTCATT 679
     |:::||||||               :::::::::|||
 262 oArgProAlaIleArgGlyAspValSerValGlyGlyMet***CysCysS 279

 680 TCATTGAGCCGGTC...GGTGCAAACAAAACCGTTTGGACCATCAATTA. 725
               ||||||   |||:::::::::   :::   :::::::::
 279 erGlyGlyProValAlaGlySerThrGlnGlyIleGly***ValGlyGly 295

 726 ...TCAAGATGTAATTGCCATCGGACGTTTGTTTGCAACAGGCCGTCTGA 772
          ||||||::::::   |||   ::::::   :::||||||
 296 HisArgArgCysSerAla...ArgGlnLeuLeuArgThrArgPro..... 309

 773 ACACCGAGCGCGTGATTGC..............TTTGGGTGGTTCTCAA 807
      ||||||   |||   |||               :::|||      :::|
 310 .HisArgArgArgArgCysArgArgGlySerArgIleGlyGlyGlyAlaS 326

 808 GTCAACAAACCACGCCTCTTGCGTACCGTTTTGGGTGCGAAAGTATCGCA 857
     ||               |||:::   ||||||      :::   ||||||
 326 er***ProAspArgAspLeuGlyAlaArgPheArgAspGlnArgIleAla 342

 858 AATTACTGC..............GGGCGAATTGGTTGACGCAGACAACC 892
     :::                     |||  |||   ||||||:::
 343 GlyArgTrpLeuAspAlaGlyProArgArgAlaGlyGlyArgArgArgAr 359

 893 GCGTGATTTCCGG.............................TTCG 909
      |||      |||                             |
 359 gArgCysArgArgAlaValArgArgGlyGlyArgLeuArgAlaAlaThrG 376

 910 GTATTGAACGGCGCGATTACACAAGG..................... 935
     ||   :::|||||||||      |||
 376 lySerArgArgArgAspThrGlyArgArgTyrGlnCysProProAlaGly 392

 935 ..................................................... 935

 393 AlaGlyArgGlyArgHisArgArgArgAlaArgAspGlyAlaAlaGlnTr 409

 936 ..........CGCGCACGATTATTTGGGACGCTACCACAATCAGATTTCC 975
               |||   |||      |||   :::   :::::::|||   |
 409 pLeuCysGlyArgArgArgThrGlyGlyArgValTyrArgGlyAsp...A 425

 976 GTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCGCA 1025
     ||      ||||||          ||||||   |||:::      |||:::
 425 rgLeuGlyArgArgArgGlyThrArgAlaAspArgIleAspGlyAlaGly 441

1026 GCCGGACAAATA..................................... 1037
     :::|||:::
 442 ValGlyArgAlaGlyArgAla***ArgGlyProProGlyArgArgArgAr 458

1038 .CTCCATCACGCGTACGACCCTCGGCCATTTCCTGAAAAACAAACTCTTC 1086
      |||:::|||:::          |||   ||||||   :::         :
 458 gLeuGlnHisGlyProCysArgArgCysPheProAlaArgIleGlyAlaH 475

1087 AAGTTCACGACAGC..................................... 1100
     ::   |||   :::
 475 isCysHisProLysGlyAlaAspLeuGlyArgTyrLeuGlnAlaGlyArg 491

1101 CGTCAACGGTGGCGACCGCGCCATGGTGCCGATTGGTACTTACGAGCGCG 1150
     |||:::   :::|||   |||:::|||||||||
 492 ArgSerGlyTyrArgGlyArgArgGlyAlaAsp.................. 502

1151 TAATGCCGCTAGACATCCTGCCTACCCTGCTTTTGCGCGATTTAATCGTC 1200
                                                  :::||||
 503 .....................................ArgArgA 505

1201 GGCGATACCGACAG.........CGCGCAAGCATTGGGTTGCTTGGAATT 1241
     |||||   |||:::          |||::::::      ::::::||||||
 505 rgArgCysArgArgGlyGlyHisArgSerGlyArgProValValGlyIle 521

1242 GGACGAAGA 1250
     |||||||||
 522 GlyArgArg 524
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW89449

seq_documentation_block:

```
ID   AAW89449 standard; Protein; 372 AA.
XX
AC   AAW89449;
XX
DT   18-MAR-1999  (first entry)
XX
DE   A gidA2 polypeptide fragment.
XX
KW   gidA2; Staphylococcus aureus WCUH29; bacterial infection;
KW   Helicobacter pylori infection; cancer; ulcer; gastritis; antibacterial;
KW   wound treatment; bacterial adhesion; matrix protein.
XX
OS   Staphylococcus aureus.
XX
PN   EP889131-A2.
XX
PD   07-JAN-1999.
XX
PF   30-JUN-1998;   98EP-0305203.
XX
PR   01-JUL-1997;   97US-0051380.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Burnham M,  Debouck CM,  Kallender H,  Lenox AL,  Mooney JL;
PI   Palmer LM,  Zhong YY;
XX
DR   WPI; 1999-062662/06.
DR   N-PSDB; AAV82085.
XX
PT   New isolated gidA2 polypeptide from Staphylococcus aureus - used to
PT   diagnose, treat and prevent bacterial infections e.g. S. aureus and
PT   H. pylori, related cancers, ulcers and gastritis and to prevent
PT   adhesion of bacteria to matrix proteins
XX
PS   Claim 1; Page 6; 41pp; English.
XX
CC   The present sequence represents a gidA2 protein fragment of
CC   Staphylococcus aureus WCUH29. The gidA2 proteins and nucleic acids
CC   are used to treat conditions requiring increased activity or expression
```

```
CC   of gidA2, while conditions (e.g. bacterial infections) requiring
CC   inhibition of such activity or expression are treated by administering
CC   an antagonist, inhibitory nucleic acid or competitive polypeptide.
CC   Infection by S. aureus is treated, but also Helicobacter pylori
CC   infections and related cancers, ulcers and gastritis. These antibacterial
CC   agents may also be used to treat in-dwelling devices to prevent infection
CC   or generally as wound treatments to prevent adhesion of bacteria to
CC   matrix proteins.
XX
SQ   Sequence   372 AA;


alignment_scores:
            Quality:  105.00              Length:    272
              Ratio:  0.784                Gaps:     13
 Percent Similarity:  49.265   Percent Identity: 21.324

alignment_block:
 US-09-303-518D-127 x AAW89449   ..

 Align seg 1/1  to: AAW89449  from: 1  to: 372

     331 GCAAACTTAAGCGGCGANGAANTNNGNNGCAATCTGATCCAATCCGGTTT 380
         |||:::|||:::|||:::|||::::::   :::|||   :::   |||::
      10 AlaGlyLeuAlaGlySerGluAlaAlaTyrGlnLeuAlaGluArgGlyIl 26

     381 GTGGACTGCGCTG...CGTANCCGTCCGTTCAGCAAAATCCCTGCCGTCG 427
         :   :::   |||   :::   ||||||   ::::::   ||||||
      26 eLysValAsnLeuIleGluMetArgProValLysGlnThrProAlaHisH 43

     428 ATGCCGAGCCGTTCGCCATCTTCGTC......AATGCGATGGACACCAAT 471
           ::::::   ||||||   :::|||      |||::::::      |||
      43 isThrAspLysPheAlaGluLeuValCysSerAsnSerLeuArgGlyAsn 59

     472 CCGCTNGCGGCAGACCCTGTGGTTGTGATCAAAGAAGCCGNCGANGATTT 521
         |||   ::::::   |||   |||:::||||||         :::::
      60 AlaLeu...ThrAsnGlyValGlyValLeuLysGlu.........GluMe 72

     522 CAGACGANGTNTGCTGGTATTGAGCCGTTTGACCGAGCGTAAAATCCATG 571
         :||||||         ::::::   :::   :::::::::
      72 tArgArgLeuAsnSerIleIleIleGluAlaAlaAspLys.........85

     572 TGTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGCTGCCAACATC 621
                         |||   ||||||:::   :::|||      :::
      86 .................AlaArgValProAlaGlyGlyAlaLeuAlaVal 96

     622 GAAACACATGAATTCGGCGGCCCGCATCCGGCCGGTTTGAGTGGCACGCA 671
         :::   |||:::|||:::|||
      97 AspArgHisAspPheSerGly............................ 103

     672 CATTCATTTCATTGAGCCGGTCGGTGCAAACAAAACCGTTTGGACCATCA 721
             ::::::   |||   :::      ::::::::::::   :::||||
     104 ....TyrIleThrGluThrLeuLysAsnHisGluAsnIleThrValIleA 119

     722 ATTATCAAGATGTAATTGCCATCGGA.........CGTTTGTTTGCAACA 762
         ||   :::::::::   ||||||               ::::::||||||
     119 sn...GluGluIleAsnAlaIleProAspGlyTyrThrIleIleAlaThr 134

     763 GGCCGTCTGAACACCGAG............CGCGTGATTGCTTTGGGTGG 800
         |||   |||:::||||||            :::::::::   :::   ||
     135 GlyProLeuThrThrGluThrLeuAlaGlnGluIleValAspIleThrGl 151

     801 TTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTTGGGTGCGAAAG 850
         |::::::                  :::      :::      :
     151 yLysAsp....................GlnLeuTyrPheTyrAspAlaA 161

     851 TATCGCAAATTACTGCGGGCGAATTGGTTGACGCAGACAACCGCGTGATT 900
         :::::   |||         |||   :::|||   |||:::      :::
     161 laAlaProIleIleGluLysGluSerIleAspMetAspLysValTyrLeu 177

     901 ....................TCCGGTTCGGTATTGAACGGCGCGATTAC 929
                             :::::::::   ||||||      :::||
     178 LysSerArgTyrAspLysGlyGluAlaAlaTyrLeuAsnCysProMetTh 194

     930 ACAAGGCGCGCACGATTATTTGGGACGCTACCACAATCAGATT..... 972
         |:::         |||   ::::::|||:::::::::   :::
     194 rGlu.........AspGluPheAsnArgPheTyrAspAlaValLeuGluA 208

     972 ............................................... 972

     208 laGluValAlaProValAsnSerPheGluLysGluLysTyrPheGluGly 224

     973 ..............TCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTT 1007
                           :::   |||:::||||||:::      |||||
     225 CysMetProPheGluValMetAlaGluArgGlyArgLysThrLeuLeuPh 241

    1008 CGGCTGGGTTGCGCCG 1023
         ||||   :::   |||
     241 eGlyProMetLysPro 246

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAW89448

seq_documentation_block:
ID   AAW89448 standard; Protein; 435 AA.
XX
AC   AAW89448;
XX
DT   18-MAR-1999  (first entry)
XX
DE   A gidA2 polypeptide sequence.
XX
KW   gidA2; Staphylococcus aureus WCUH29; bacterial infection;
KW   Helicobacter pylori infection; cancer; ulcer; gastritis; antibacterial;
KW   wound treatment; bacterial adhesion; matrix protein.
XX
OS   Staphylococcus aureus.
XX
PN   EP889131-A2.
XX
PD   07-JAN-1999.
XX
PF   30-JUN-1998;   98EP-0305203.
XX
PR   01-JUL-1997;   97US-0051380.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Burnham M,  Debouck CM,  Kallender H,  Lenox AL,  Mooney JL;
PI   Palmer LM,  Zhong YY;
XX
DR   WPI; 1999-062662/06.
DR   N-PSDB; AAV82084.
XX
PT   New isolated gidA2 polypeptide from Staphylococcus aureus - used to
PT   diagnose, treat and prevent bacterial infections e.g. S. aureus and
PT   H. pylori, related cancers, ulcers and gastritis and to prevent
PT   adhesion of bacteria to matrix proteins
XX
PS   Claim 1; Page 4-5; 41pp; English.
XX
CC   The present sequence represents a gidA2 polypeptide of Staphylococcus
CC   aureus WCUH29. The gidA2 proteins and nucleic acids are used to treat
CC   conditions requiring increased activity or expression of gidA2, while
CC   conditions (e.g. bacterial infections) requiring inhibition of such
CC   activity or expression are treated by administering an antagonist,
CC   inhibitory nucleic acid or competitive polypeptide. Infection by
CC   S. aureus is treated, but also Helicobacter pylori infections and
CC   related cancers, ulcers and gastritis. These antibacterial agents may
CC   also be used to treat in-dwelling devices to prevent infection or
CC   generally as wound treatments to prevent adhesion of bacteria to matrix
CC   proteins.
XX
SQ   Sequence   435 AA;
```

```
alignment_scores:
            Quality:  105.00              Length:    272
              Ratio:  0.784                Gaps:     13
 Percent Similarity:  49.265   Percent Identity: 21.324

alignment_block:
 US-09-303-518D-127 x AAW89448   ..

 Align seg 1/1  to: AAW89448  from: 1  to: 435

     331 GCAAACTTAAGCGGCGANGAANTNNGNNGCAATCTGATCCAATCCGGTTT 380
         |||:::|||:::|||:::|||:::::::   :::|||   :::   |||::
      10 AlaGlyLeuAlaGlySerGluAlaAlaTyrGlnLeuAlaGluArgGlyIl 26

     381 GTGGACTGCGCTG...CGTANCCGTCCGTTCAGCAAAATCCCTGCCGTCG 427
         :   :::   |||   :::   ||||||   ::::::   ||||||
      26 eLysValAsnLeuIleGluMetArgProValLysGlnThrProAlaHisH 43

     428 ATGCCGAGCCGTTCGCCATCTTCGTC......AATGCGATGGACACCAAT 471
         ::::::   ||||||   :::|||      |||:::::      |||
      43 isThrAspLysPheAlaGluLeuValCysSerAsnSerLeuArgGlyAsn 59

     472 CCGCTNGCGGCAGACCCTGTGGTTGTGATCAAAGAAGCCGNCGANGATTT 521
            |||   ::::::   |||   |||:::||||||         :::::
      60 AlaLeu...ThrAsnGlyValGlyValLeuLysGlu.....GluMe 72

     522 CAGACGANGTNTGCTGGTATTGAGCCGTTTGACCGAGCGTAAAATCCATG 571
         :||||||         ::::::   :::   :::::::::
      72 tArgArgLeuAsnSerIleIleIleGluAlaAlaAspLys..... 85

     572 TGTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGCTGCCAACATC 621
                        |||   ||||||:::   :::|||      :::
      86 .................AlaArgValProAlaGlyGlyAlaLeuAlaVal 96

     622 GAAACACATGAATTCGGCGGCCCGCATCCGGCCGGTTTGAGTGGCACGCA 671
         :::   |||:::|||:::|||
      97 AspArgHisAspPheSerGly............................. 103

     672 CATTCATTTCATTGAGCCGGTCGGTGCAAACAAAACCGTTTGGACCATCA 721
             ::::::   |||   :::      ::::::::::::   :::|||||
     104 ....TyrIleThrGluThrLeuLysAsnHisGluAsnIleThrValIleA 119

     722 ATTATCAAGATGTAATTGCCATCGGA.........CGTTTGTTTGCAACA 762
         ||   :::::::::   ||||||               ::::::||||||
     119 sn...GluGluIleAsnAlaIleProAspGlyTyrThrIleIleAlaThr 134

     763 GGCCGTCTGAACACCGAG............CGCGTGATTGCTTTGGGTGG 800
         |||   |||:::||||||            :::::::::   :::   ||
     135 GlyProLeuThrThrGluThrLeuAlaGlnGluIleValAspIleThrGl 151

     801 TTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTTGGGTGCGAAAG 850
         |::::::                  :::      :::      :
     151 yLysAsp.....................GlnLeuTyrPheTyrAspAlaA 161

     851 TATCGCAAATTACTGCGGGCGAATTGGTTGACGCAGACAACCGCGTGATT 900
         :::::   |||         |||   :::|||   |||:::      :::
     161 laAlaProIleIleGluLysGluSerIleAspMetAspLysValTyrLeu 177

     901 ....................TCCGGTTCGGTATTGAACGGCGCGATTAC 929
                              :::::::::   ||||||      :::||
     178 LysSerArgTyrAspLysGlyGluAlaAlaTyrLeuAsnCysProMetTh 194

     930 ACAAGGCGCGCACGATTATTTGGGACGCTACCACAATCAGATT....... 972
         |:::         |||   ::::::|||:::::::::   :::
     194 rGlu.........AspGluPheAsnArgPheTyrAspAlaValLeuGluA 208

     972 ................................................. 972

     208 laGluValAlaProValAsnSerPheGluLysGluLysTyrPheGluGly 224

     973 ..............TCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTT 1007
                        :::   |||:::||||||:::      |||||
     225 CysMetProPheGluValMetAlaGluArgGlyArgLysThrLeuLeuPh 241

     1008 CGGCTGGGTTGCGCCG 1023
          ||||   :::   |||
      241 eGlyProMetLysPro 246

 seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW13502

 seq_documentation_block:
 ID   AAW13502 standard; protein; 1645 AA.
 XX
 AC   AAW13502;
 XX
 DT   30-JAN-1998  (first entry)
 XX
 DE   B. pertussis adenylcyclase-haemolysin mutant delta-Cla.
 XX
 KW   Mutant; Bordetella pertussis; adenylcyclase; haemolysin; wild type;
 KW   deletion; induction; antibody; Bordetella parapertussis; vaccine; human;
 KW   Bordetella bronchiseptica; veterinary.
 XX
 OS   Bordetella pertussis.
 OS   Synthetic.
 XX
 FH   Key             Location/Qualifiers
 FT   Region          826..827
 FT                   /note= "deletion of amino acids 827-887 from the wild
 FT                          type sequence between these positions"
 FT   Modified-site   919..924
 FT                   /note= "Lys at position 922 is modified by addition of
 FT                          a fatty acid"
 XX
 PN   FR2736064-A1.
 XX
 PD   03-JAN-1997.
 XX
 PF   30-JUN-1995;   95FR-0007945.
 XX
 PR   30-JUN-1995;   95FR-0007945.
 XX
 PA   (INSP ) INST PASTEUR.
 XX
 PI   Betsou F,  Guiso N,  Sebo P;
 XX
 DR   WPI; 1997-111454/11.
 XX
 PT   New immunogenic fragments of Bordetella adenyl:cyclase haemolysin -
 PT   and related nucleic acid and antibodies, for use in vaccines and
 PT   immunotherapy
 XX
 PS   Claim 6; Page -; 55pp; French.
 XX
 CC   This is the sequence of a novel mutant, designated delta-Cla, derived
 CC   from the Bordetella pertussis adenylcyclase-haemolysin (AC-Hly) in which
 CC   amino acids 827-887 of the wild type sequence have been deleted by
 CC   genetic engineering.  The novel protein is able to induce production
 CC   of protective antibodies against the Bordetella species pertussis,
 CC   parapertussis and/or bronchiseptica, especially in protective vaccines
 CC   for human or veterinary use.
 CC   Note: this sequence is not given in the specification but is generated
 CC   from the wild type sequence.
 XX
 SQ   Sequence   1645 AA;


alignment_scores:
            Quality:  104.50              Length:    455
              Ratio:  0.522                Gaps:     24
 Percent Similarity:  43.956   Percent Identity: 21.538

alignment_block:
 US-09-303-518D-127 x AAW13502   ..

 Align seg 1/1  to: AAW13502  from: 1  to: 1645
```

```
190 CCGGGCGTGGTGTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCATCCA 239
694 ProValAlaValValThrSerLeuLeuThrGlyAlaLeuAsnGlyIleLe 710

240 TCGCGGCGAAAAGCGCGTACTTCAGTCGGTCGTGATTGCCGTTGAAGGCA 289
710 uArgGly.....ValGlnGlnProIleIleGluLysLeuAlaA 723

290 ACGAC.....GAAATCGAGTTC 306
723 snAspTyrAlaArgLysIleAspGluLeuGlyGlyProGlnAlaTyrPhe 739

307 GAACGCTACGCGCCC.....GAAGCGTTGGCAAACTTAAGCGGCGA 347
740 GluLysAsnLeuGlnAlaArgHisGluGlnLeuAlaAsnSerAspGlyLe 756

348 NGAANTNNGNNGCAATCTGATCCAATCCGGTTTGTGGACTGCG..... 390
756 uArgLysMetLeuAlaAspLeuGlnAlaGly...TrpAsnAlaSerSerV 772

391 .....CTGCGTANCCGTCCGTTCAGCAAA..... 414
772 alIleGlyValGlnThrThrGluIleSerLysSerAlaLeuGluLeuAla 788

415 .....ATCCCTGCCGTCGATGCCGAGCCGTT 440
789 AlaIleThrGlyAsnAlaAspAsnLeuLysSerValAsp..... 801

441 CGCCATCTTCGTCAATGCG...ATGGACACCAATCCGCTNGCGGCAGACC 487
802 ....ValPheValAspArgPheValGlnGlyGluArgValAlaGlyGlnP 817

488 CTGTGGTTGTGATCAAAGAAGCCGNCGANGATTTCAGACGANGTNTGCTG 537
817 roValValLeuAspValAlaAlaGlyGlyAspLeuAlaLys..... 830

538 GTATTGAGCCGTTTGACCGAGCGT.....AA 563
831 ValValSerGlnLeuValAspAlaAsnGlyValLeuLysHisSerIleLy 847

564 AATCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGCTG 613
847 sLeuAspValIleGlyGlyAspGlyAspAspValValLeuAlaAsnAlaS 864

614 CCAACATCGAAACACATGAATTCGGCGGCCCGCATCCGGCCGGTTTGAGT 663
864 erArgIle.....HisTyrAspGlyGlyAla 872

664 GGCACGCACATTCATTTCATTGAGCCGGTCGGTGCAAACAAAACCGTTTG 713
873 GlyThrAsn..... 875

714 GACCATCAATTAT.....CAAGATGTAATTGCCATCGGAC 748
876 .ThrValSerTyrAlaAlaLeuGlyArgGlnAspSerIleThrValSer. 891

749 GTTTGTTTGCAACAGGC...CGTCTGAACACCGAGCGCGTGATTGCTTTG 795
892 .....AlaAspGlyGluArgPheAsnValArgLys..... 901

796 GGTGGTTCTCAAGTCAACAAACCACGCCTCTTGCGT..... 831
902 .....GlnLeuAsnAsnAlaAsnValTyrArgGluGlyValAlaTh 915

832 .....ACCGTTTTGGGTGCGAAAGTATCGCAAATT..... 861
915 rGlnThrThrAlaTyrGlyLysArgThrGluAsnValGlnTyrArgHisV. 932

862 .....ACTGCGGGCGAATTGGTTGACGCAGACAAC..... 891
932 alGluLeuAlaArgValGlyGlnValValGluValAspThrLeuGluHis 948

892 ...CGCGTGATTTCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGC 938
949 ValGlnHisIleIleGlyGlyAlaGlyAsnAspSerIleThrGlyAsnAl 965

939 GCACGATTATTTG.....GGAC 955
965 aHisAspAsnPheLeuAlaGlyGlySerGlyAspAspArgLeuAspGlyG 982

956 GCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGC..... 993
982 lyAlaGlyAsnAspThrLeuValGlyGlyGluGlyGlnAsnThrValIle 998

994 .....AGCAA 998
999 GlyGlyAlaGlyAspAspValPheLeuGlnAspLeuGlyValTrpSerAs 1015

999 AGAGCTGTTCGGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGC 1048
1015 nGlnLeuAspGlyGlyAlaGlyValAspThrValLysTyrAsnValHisG 1032

1049 GTACGACCCTCGGCCATTTCCTGAAAAACAAACTCTTCAAGTTCACGACA 1098
1032 lnProSer.....GluGluArgLeuGluArgMetGly... 1042

1099 GCCGTCAACGGTGGCGACCGCGCCATGGTGCCGATTGGTACTTACGAGCG 1148
1043 ......AspThrGlyIleHisAlaAspLeuGlnLysGlyThrValGluLy 1057

1149 C.....GTAATGCCGCTAGACA 1165
1057 sTrpProAlaLeuAsnLeuPheSerValAspHisValLysAsnIleGluA 1074

1166 TCCTGCCTACCCTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGC 1215
1074 snLeuHisGlySerArgLeuAsnAspArgIleAlaGlyAspAspGlnAsp 1090

1216 GCGCAAGCATTGGGT 1230
1091 AsnGluLeuTrpGly 1095
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:AAP93285

seq_documentation_block:
ID AAP93285 standard; protein; 3080 AA.
XX
AC AAP93285;
XX
DT 17-DEC-2001 (updated)
DT 06-APR-1990 (first entry)
XX
DE Sequence of clone HIV-2 SBL/ISY.
XX
KW HIV-2; proviral clone HIV-2 SBL/ISY;
XX
OS Human immunodeficiency virus 2.
XX
PN USN7331212-N.
XX
PD 29-AUG-1989.
XX
PF 31-MAR-1989; 89US-0331212.
XX
PR 31-MAR-1989; 89US-0331212.
XX
PA (USSH) US DEPT. HEALTH AND HUMAN SERVICES.
XX
PI Franchini G, Wong-Staal F, Gallo R;
XX
DR WPI; 1989-339698/46.
DR N-PSDB; AAN92119.
XX
PT Complete human immunodeficiency type 2 proviral clone - used to generate
PT animal model for function studies of HIV genes in vivo.

```
XX
PS   Disclosure; Fig. 5; 43pp; English.
XX
CC   The protein is encoded by the third reading frame of HIV-2 SBL/ISY, a
CC   proviral clone of HIV-2.
CC   (Note: Revised entry submitted to correct the patent number format of
CC   US Government-owned NTIS applications to prevent clashes with ongoing US
CC   granted patent numbers. For further information please visit the Derwent
CC   web site at www.derwent.com/dwpi/updates/ntis_us.html.)
XX
SQ   Sequence   3080 AA;


alignment_scores:
           Quality:  104.50              Length:    433
             Ratio:  0.562                Gaps:     25
 Percent Similarity:  42.956   Percent Identity: 21.940

alignment_block:
 US-09-303-518D-127 x AAP93285   ..

 Align seg 1/1  to: AAP93285  from: 1  to: 3080

     195 CGTGGTGTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCATCCATCGCG 244
         |||   :::      |||      |||:::::::::::::::   :::
     256 ArgHisLeuValLeuAlaArgArgArgGluSerGluArgTyrArgSerLy 272

     245 GCGAAAAGCGCGT...ACTTCAGTCGGT......CGTGATTGCCGTTGAA 285
         ::::::::::|||   ||||||   |||      |||:::||||||   :
     272 sLysAsnSerArgGluThrSerSerGlyArgAsnArgAsnCysArgGluA 289

     286 GGCAACGACGAAATCGAGTTCGAACGCTACGCGCCCGAAGCGTTGGCAAA 335
         ::   :::   :::          |||   |||   |||      ||||||
     289 snAlaLysTyrLysThrAsnSerThrThrArgGluArg......GlyLys 303

     336 CTTAAGCGGCGANGAANTNNGNNGCAATCTGATCCAATCCGGTTTGTGGA 385
         |||   |||   :::          |||
     304 LeuProArgAlaThrAsnArgArgGln..................... 312

     386 CTGCGCTGCGTANCCGTCCGTTCAGCAAAATCCCTGCCGTCGATGCCGAG 435
                 ||||||:::   ::::::|||      :::|||   :   :
     313 .......LeuCysProSerAlaAlaGluSerProAsnProLysCysLeuG 327

     436 CC...GTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTNGC... 479
         ::   |||   :::   ::::::      |||::::::||||||
     327 lyLysValSerArgGlyGluLysValArgGlyArgSerSerAlaGlyIle 343

     480 ..............................................GG 481
                                                       ||
     344 SerGlyThrLeuArgArgLeuHisAlaLeuTyrSerAsnAlaLeuCysGl 360

     482 CAGACCCTGTGGTTGTGATCAAAGAAGCCGNCGANGATTTCAGACGANGT 531
         |||||||   :::   |||   :::::::::|||   ::::::::::::
     360 yArgProSerSerSerAspAlaAsnAsnGlnArgAsnTyrArgSerSerA 377

     532 NTGCTGGTATTGAGCCGTTTGACCGAGCGTAAAATCCATGTGTGTAAGGC 581
             |||
     377 rgLeuGly.......................................... 379

     582 AGCTGGCGCAGACGTGCCGTCTGAAAATGCTGCCAACATCGAAACACATG 631
                                     |||         ||||||
     380 ............................CysThrThrSerAsnThr.. 385

     632 AATTCGGCGGCCCGCATCCGGCCGGTTTGAGT......GGCACGCACATT 675
             ||||||   ::: :::|||   :::      :::|||:::|||
     386 .......ArgProLeuThr.SerGlyAlaAlaGlnArgSerThrArgIle 399

     676 CATTTCATTGAGCCGGTCGGTGCAAACAAAACCGTTTGGACCATCAATTA 725
         |||      :::         ::::::  |||
     400 HisSerArgAspAsnLysHisSerArgGlyThr................. 410

     726 TCAAGATGTAATTGCCATCGGACGTTTGTTTGCAACAGGCCGTCTGAACA 775
             |||   :::   ::::::|||      :::|||:::|||   :::

     411 ....AspArgMetAspValAlaArgLysSerCysThrSerArgLysHisL 426

     776 CCGAG.............................CGCGTGATTGCTTTG 795
           |||                             :::   :::
     426 euGluMetAspProAspArgThrAlaGluValCysGlnAsnValGlnSer 442

     796 GGTGGTTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTT........ 837
         :::      :::   :::|||   |||      |||:::|||
     443 AsnGlnHisSerArgHisLysThrArgThrLysArgValValProLysLe 459

     838 ..................TTGGGTGCGAAAGTATCGCAAATTACTGCGG 868
                                |||   ::::::::::::::   ::::::
     459 uCysGlyIleLeuGlnLysLeuLysGlyArgThrAspArgCysSerSerG 476

     869 GCGAATTGGTTGACGCAGAC.................AACCGCGTGATT 900
           ||||||   |||   |||                        |||::::::
     476 luGluLeuAspAspProAspAlaAlaSerAlaIleGluProArgLeuVal 492

     901 TCCGGTTCGGTATTGAACGGCGCGATTACACAAGGCGCGCACGAT..... 945
         |||                        |||:::|||::::::|||
     493 Ser.......................ThrLysGlyThrArgAspGluSe 501

     946 .TATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAGAAGGCCGCA 994
          ||||||   |||      |||::::::|||                |||:
     501 rTyrLeuArgArgAspAlaAsnArgLeuSer..............ArgA 513

     995 GCAAAGAGCTGTTCGGCTGGGTTGCGCCGCAGCC.........GGACAA 1034
         :::::            |||:::   |||:::            |||:::
     513 spArg............TrpThrArgProGluGlyGlnThrAsnGlyArg 525

    1035 ATACTCCATCACGCGTACGACCCTCGGCCATTTCCT.............. 1070
           |||   :::::::::          |||:::|||
     526 SerLeuLysArgGlyHisAlaThrSerProTyrProIleCysSerSerPr 542

    1071 ....GAAAAA...CAAACTCTTCAAGTTCACGACAGCCGT.......... 1103
           ||||||   ::::::|||::::::      :::|||
     542 oThrGluLysGlyAsnValLeuGluLeuTrpLysGlyArgAlaLeuGlyL 559

    1104 ..............................CAACGGTGG......CGA 1115
                                        :::   |||      |||
     559 ysThrMetProSerAlaLysThrGlyLeuLeuGluMetTrpGlnValArg 575

    1116 CCGCGCCATGGT............GCCGATTGGTACTTACGAGCG..... 1148
           :::|||||           :::   |||::::::||||||
     576 ThrHisHisGlyLysLeuProArgThrGlyTrpPhePheArgAlaTrpTh 592

    1149 .......CGTAATGCCGCTAGACATCCTGCCTACCCTGCTTTTGCGCGAT 1191
               :::::::|||   :::   |||      |||   ||||||
     592 rMetGlyLysGluAlaProGlnLeuProArgGlyProLysPheAlaGlyA 609

    1192 TTAATCGTCGGCGATACCGACAGCGCGCAAGCATTGGGTTGCTTGGA 1238
           |||   :::         :::   :::|||   |||      |||
     609 laAsnThrAsnSerThrProAsnGlySerSerSerGlyProThrGly 624

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU34397

seq_documentation_block:
ID   AAU34397 standard; Protein; 433 AA.
XX
AC   AAU34397;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #673.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
```

```
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS52256.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 5893; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   433 AA;


alignment_scores:
            Quality:  104.00              Length:    272
              Ratio:  0.776                Gaps:     13
 Percent Similarity:  49.265   Percent Identity: 21.324

alignment_block:
 US-09-303-518D-127 x AAU34397   ..

 Align seg 1/1  to: AAU34397  from: 1  to: 433

     331 GCAAACTTAAGCGGCGANGAANTNNGNNGCAATCTGATCCAATCCGGTTT 380
         |||:::|||:::|||:::|||::::::   :::|||   :::   |||::
       8 AlaGlyLeuAlaGlySerGluAlaAlaTyrGlnLeuAlaGluArgGlyIl 24

     381 GTGGACTGCGCTG...CGTANCCGTCCGTTCAGCAAAATCCCTGCCGTCG 427
         :   :::   |||   :::   ||||||   ::::::   ||||||
      24 eLysValAsnLeuIleGluMetArgProValLysGlnThrProAlaHisH 41

     428 ATGCCGAGCCGTTCGCCATCTTCGTC......AATGCGATGGACACCAAT 471
           ::::::   ||||||   :::|||      |||::::::      |||
       41 isThrAspLysPheAlaGluLeuValCysSerAsnSerLeuArgGlyAsn 57

     472 CCGCTNGCGGCAGACCCTGTGGTTGTGATCAAAGAAGCCGNCGANGATTT 521
           |||   ::::::   |||   |||:::||||||         :::::
       58 AlaLeu...ThrAsnGlyValGlyValLeuLysGlu.........GluMe 70

     522 CAGACGANGTNTGCTGGTATTGAGCCGTTTGACCGAGCGTAAAATCCATG 571
         :||||||         ::::::   :::   :::::::::
      70 tArgArgLeuAsnSerIleIleIleGluAlaAlaAspLys..... 83

     572 TGTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGCTGCCAACATC 621
                         |||   ||||||:::   :::|||      :::
      84 .................AlaArgValProAlaGlyGlyAlaLeuAlaVal 94

     622 GAAACACATGAATTCGGCGGCCCGCATCCGGCCGGTTTGAGTGGCACGCA 671
         :::   |||:::|||:::|||
      95 AspArgHisAspPheSerGly............................. 101

     672 CATTCATTTCATTGAGCCGGTCGGTGCAAACAAAACCGTTTGGACCATCA 721
           ::::::   |||   :::      ::::::::::::   :::||||
     102 ....TyrIleThrGluAsnLeuLysAsnHisGluAsnIleThrValIleA 117

     722 ATTATCAAGATGTAATTGCCATCGGA.........CGTTTGTTTGCAACA 762
         ||   :::::::::   ||||||               ::::::||||||
     117 sn...GluGluIleAsnAlaIleProAspGlyTyrThrIleIleAlaThr 132

     763 GGCCGTCTGAACACCGAG............CGCGTGATTGCTTTGGGTGG 800
         |||   |||:::||||||            :::::::::   :::   ||
     133 GlyProLeuThrThrGluThrLeuAlaGlnGluIleValAspIleThrGl 149

     801 TTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTTGGGTGCGAAAG 850
         |::::::                  :::      :::      :
     149 yLysAsp.....................GlnLeuTyrPheTyrAspAlaA 159

     851 TATCGCAAATTACTGCGGGCGAATTGGTTGACGCAGACAACCGCGTGATT 900
         :::::   |||         |||   :::|||   |||:::      :::
     159 laAlaProIleIleGluLysGluSerIleAspMetAspLysValTyrLeu 175

     901 .....................TCCGGTTCGGTATTGAACGGCGCGATTAC 929
                               :::::::::   ||||||      :::||
     176 LysSerArgTyrAspLysGlyGluAlaAlaTyrLeuAsnCysProMetTh 192

     930 ACAAGGCGCGCACGATTATTTGGGACGCTACCACAATCAGATT....... 972
         |:::         |||   ::::::|||::::::::   :::
     192 rGlu.........AspGluPheAsnArgPheTyrAspAlaValLeuGluA 206

     972 ..................................................... 972

     206 laGluValAlaProValAsnSerPheGluLysGluLysTyrPheGluGly 222

     973 ..............TCCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTT 1007
                           :::   |||:::|||||||:::      |||||
     223 CysMetProPheGluValMetAlaGluArgGlyArgLysThrLeuLeuPh 239

    1008 CGGCTGGGTTGCGCCG 1023
         ||||   :::   |||
     239 eGlyProMetLysPro 244

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW18663

seq_documentation_block:
ID   AAW18663 standard; Protein; 387 AA.
XX
AC   AAW18663;
XX
DT   24-JUL-1997  (first entry)
XX
DE   Fragmented human NF-H gene +2 frameshift mutant product.
XX
KW   Frameshift mutation product; GAGA motif; somatic mutation; diagnosis;
KW   detection; antibody; probe; cancer; neoplasia; neurodegenerative;
KW   Parkinson's; Alzheimer's disease; Pick's; Huntington's disease;
KW   Down's syndrome; frontal lobe dementia; progressive supranuclear palsy;
KW   PSP; amyotrophic lateral sclerosis; multiple sclerosis; MS;
KW   cardiovascular; rheumatoid arthritis.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
```

```
FT   Misc-difference 1..387
FT                   /note= "X corresponds to a stop codon in the
FT                          accompanying DNA file, AAT69796"
XX
PN   WO9712992-A2.
XX
PD   10-APR-1997.
XX
PF   02-OCT-1996;   96WO-IB01106.
XX
PR   11-JAN-1996;   96US-0009832.
PR   02-OCT-1995;   95GB-0020080.
XX
PA   (ROYA-) ROYAL NETHERLANDS ACAD ARTS & SCI.
PA   (UYRO-) UNIV ROTTERDAM ERASMUS.
PA   (UYUT-) UNIV STATE UTRECHT.
XX
PI   Burbach JPH,  Grosveld FG,  Van Leeuwen FW;
XX
DR   WPI; 1997-226235/20.
DR   N-PSDB; AAT69795.
XX
PT   Use of mutant genes having frame:shift mutation(s) - for developing
PT   prods. for the diagnosis, prevention and treatment of associated
PT   diseases, e.g. cancer or neuro:degenerative disease
XX
PS   Claim 22; Fig 9; 123pp; English.
XX
CC   AAW18663 and AAW18664 are +2 and +1 frameshift mutations, respectively,
CC   of a sequence comprising fragments of the coding sequence of the
CC   human neurofilament subunit NF-H gene corresponding to nucleotides
CC   1-1162 of the wild-type NF-H gene. This region contains GAGAG motifs.
CC   Frameshift mutants of the tau, ubiquitin, apolipoprotein E,
CC   microtubule-associated protein 2 (MAP-2), neurofilament subunit L, M
CC   and H and amyloid A4 genes are claimed. All these genes share a common
CC   GAGAN motif (N= A, G, C or T), which is the site of common GA
CC   dinucleotide deletion(s) that cause neurodegenerative disorders.
CC   Antigenic peptides used for the production of antibodies, and small
CC   nucleic acid sequences derived from frameshift mutants are used in the
CC   diagnosis, prevention and treatment of cancer and neurodegenerative
CC   diseases, e.g. Parkinson's disease, Alzheimer's disease, Down's
CC   syndrome, frontal lobe dementia (Pick's disease), progressive
CC   supranuclear palsy (PSP), amyotrophic lateral sclerosis, Huntington's
CC   disease, multiple sclerosis, and other degenerative diseases such as
CC   cardiovascular disease and rheumatoid arthritis.
XX
SQ   Sequence   387 AA;
```

```
alignment_scores:
           Quality:  103.50              Length:    455
             Ratio:  0.627                Gaps:     23
 Percent Similarity:  36.264   Percent Identity: 21.538

alignment_block:
 US-09-303-518D-127 x AAW18663   ..

 Align seg 1/1  to: AAW18663  from: 1  to: 387

      68 GGCCCGTCATTACCGAAGTCGCGTTGCTTG..................... 97
         |||||||||   |||   |||:::|||
      51 GlyProSerSerProProSerHisCysGlnGlyValGlyProGlyArgGl 67

      98 ..................GCGAAGAATATGCCGGTATGCGCCCCT.... 124
                            |||      :::|||   |||   |||
      67 yGlyTyrLysArgAlaGlyAlaLeuValValProGlnCysLeuProProA 84

     125 ................................NGATGAAAGTCAAGG 139
                                        :::      :::|||
      84 rgProGlyLeuAlaHisLeuLeuArgPro******AlaSerAlaAlaArg 100

     140 AAGGCGATGCCGTCAAAAAAGGCCAAGTGCTGTTTGAAGACAAAAAGNAT 189
                  :::   :::::::::   |||:::   :::   :::
```

```
     101 ThrArgCysTrpAlaProArgSerArgArgCysMetAlaAlaAlaAlaSe 117

     190 CCGGGCGTGGTGTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCATCCA 239
            :::             |||      ||||||         |||   :
     117 rThrThrArg***ProGluArgValAlaGlnAlaGlyArgAlaProProL 134

     240 TCGCGGCGAAAAGCGCGTACTTCAGTCGGTCGTGATTGCCGTTGAAGGCA 289
         ::|||                                  |||     |||
     134 euAla..............................ProProAlaAla 139

     290 ACGACGAAATCGAGTTCGAACGCTACGCGCCCGAAGCGTTGGCAAACTTA 339
         :::|||::::::   ::::::   :::   |||
     140 SerThrArgGlyHisGlyArgPro***AlaPro..................  150

     340 AGCGGCGANGAANTNNGNNGCAATCTGATCCAATCCGGTTTGTGGACTGC 389

     150 .................................................. 150

     390 GCTGCGTANCCGTCCGTTCAGCAAAATCCCTGCCGTCGATGCCGAGCCGT 439
                       |||   |||   |||:::   :::   |||   :::
     151 .....CysProProArgProAlaAlaSerValAlaGlnAlaProProGln 165

     440 TCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTNGCGGCAGA.... 485
         :::|||:::          |||||||| ::::::|||:::|||||
     166 AlaProThr.........ArgTrpThr.Arg***AlaThrGlyArgArgA 179

     486 ...................CCCTGTGGTTGTGATCAAAGAAGCCGNCGA 515
                             |||   :::      :::   |||   |||
     179 laAlaTrpTrpArgTrpProProHisAlaValArgArgSerSerCysArg 195

     516 NGATTTCAGACGANGTNTGCTGGTATTGAGCCGTTTGACCGAGCGTAAAA 565
               |||::::::   |||   :::                  ::
     196 Arg***ThrThrAlaSerProGlyThrSer.................Th 206

     566 TCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAATGCTGCC 615
         :   |||       ||||||||||||||   ::::::      :::   :
     206 rArgCys......GlySerTrpArgArgThrThrAla......AlaTrpA 219

     616 AACATCGAAACACATGAATTCGGCGGCCCGCATCCGGCCGGTTTGAG... 662
         ::   |||          :::||||||   :::|||:::|||
     219 rgAlaArg.........LeuArgArgCysGlySerSerArgArgAlaAla 232

     663 ......TGG............CACGCACATTCATTTCATTGAGCCGGTCG 694
               |||               |||:::|||   :::   |||:::||
     233 ProLeuTrpAlaSerCysThrSerAlaArgSerAlaArgCysAlaAlaAr 249

     695 GTGC.........AAACAAAACCGTTTGGACCATCAATTATCAAGATGTA 735
         ||||            ::::::   :::::::::      ||||||   :
     249 gCysCysAlaTrpAlaArgArgAlaValSerTyrAlaTrpSerArgSerT 266

     736 ATTGCCATCGGACGTTTGTTTGCAACAGGCCGTCTGAACACCGAGCGCGT 785
         ::|||   ||||||                                :::|||
     266 hrCysSerArgThr..............................SerArg 272

     786 GATTGCTTTGGGTGGTTCTCAAGTCAACAAACCACGCCTCTTGCGTACCG 835
            |||             :::|||      ||||||
     273 ThrCys............AlaSerAla***ThrThr.............. 280

     836 TTTTGGGTGCGAAAGTATCGCAAATTACTGCGGGCGAATTGGTTGACGCA 885

     280 .................................................. 280

     886 GACAACCGCGTGATTTCCGGTTCGGTATTGAACGGCGCGATTACACAAGG 935
                          |||   |||   |||||||||          |||
     281 ................ArgProGlySerGluArgArgProArgArgArg 291

     936 CGCGCACGATTATTTGGGACGCTACCACAATCAGATTTCCGTTATCGAAG 985
         ||||||      ::::::                              |||||
     292 ProAlaArgTrpArgAlaSer.....................ArgAr 300

     986 AAGGCCGCAGCAAAGAGCT.........................GTTC 1008
         |||||||::::::||||||                          |
     300 gArgProArgArgArgAlaTrpThrCysArgArgArgArgArgArgCysA 317
```

```
    1009 GGCTGGGTTGCGCCGCAGCCGGACAAATACTCCATCACGCGTACGACCCT 1058
         ||   ::::::|||::::::|||
     317 rgArgSerAlaAlaThrCysGlyAla..................... 325

    1059 CGGCCATTTCCTGAAAAACAAACTCTTCAAGTTCACGACAGCCGTCAACG 1108
                                                :::|||:::||
     326 ................................ThrThrArgLysAr 330

    1109 GTGGCGACCGCGCCATGGTGCCGATTGGTACTTACGAGCGCGTAATGCCG 1158
         ||||                            ||||||:::   |
     330 gTrpAlaSer.....................CysSerAlaArgSerArgA 340

    1159 CTAGACATCCTGCCTACCCTGCTTTTGCGCGATTTAATCGTCGGCGATAC 1208
         ||      |||                   :::|||||||||
     340 laProAlaPro.....................ArgArgArgArgCys 348

    1209 CGACAGCGCGCAAGCATTGGGTTGCTTGGAATTGGACGAAGAAGA...CC 1255
         |||   |||   :::          :::      |||||||||
     349 ArgProArgArgAlaThrPro***SerAlaThr***ArgArgArgCysAl 365

    1256 TCGCTTTGTGCAG 1268
          ||||||::::::
     365 aArgPheAlaArg 369
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB15935

seq_documentation_block:

ID AAB15935 standard; Protein; 740 AA.
XX
AC AAB15935;
XX
DT 05-OCT-2000 (first entry)
XX
DE E. coli proliferation associated protein sequence SEQ ID NO:292.
XX
KW Escherichia coli; E. coli; proliferation; inhibition; screening;
KW antimicrobial; bacterial growth; antisense therapy; antibacterial.
XX
OS Escherichia coli.
XX
PN WO200044906-A2.
XX
PD 03-AUG-2000.
XX
PF 27-JAN-2000; 2000WO-US02200.
XX
PR 27-JAN-1999; 99US-0117405.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2000-514822/46.
DR N-PSDB; AAA65940.
XX
PT Novel polynucleotides and polypeptides associated with microorganism
PT proliferation, used to identify inhibitors of bacterial growth and
PT proliferation, for use in antisense therapy -
XX
PS Claim 11; Page 217-219; 316pp; English.
XX
CC AAA65809 to AAA65889 and AAA66058 to AAA66138 represent nucleotide
CC sequences derived from Escherichia coli which inhibit E. coli
CC proliferation. AAA65890 to AAA66055 and AAB15886 to AAB16040 represent
CC nucleotide and protein sequences associated with E. coli proliferation.
CC AAA66056 and AAA66057 represent primers used for sequencing E. coli
CC proliferation inhibiting nucleotide inserts in an example from the
CC present invention. Methods from the present invention can be used to
CC identify a proliferation- required gene in a microorganism, by contacting
CC a microorganism with a proliferation-required gene activity inhibitory
CC nucleic acid identified in another organism, and determining if
CC inhibition occurs in the second microorganism. The nucleic acid sequences
CC identified as being required for bacterial growth and proliferation, can
CC be used for antisense therapy for killing bacteria.
XX
SQ Sequence 740 AA;

alignment_scores:

```
           Quality:  103.00              Length:    324
             Ratio:  0.665                 Gaps:     13
 Percent Similarity:  47.840   Percent Identity: 21.914
```

alignment_block:

US-09-303-518D-127 x AAB15935 ..

Align seg 1/1 to: AAB15935 from: 1 to: 740

```
    100 GAAGAATATGCCGGTATGCGCCCCTNGATGAAAGTCAAGGAAGGCGATGC 149
        :::::::::   |||   :::      :::   |||:::   ||||||
     48 LysGlnHisIleGlyAlaGluGlyGluLeuCysValSerValGlyAspLy 64

    150 CGTCAAAAAAGGCCAAGTGCTGTTT......GAAGACAAAAAGNATCCGG 193
         |||   :::|||||||   |||          :::   |||      |||
     64 sValLeuArgGlyGlnProLeuThrArgGlyArgGlyLysMetLeuPro. 80

    194 GCGTGGTGTTTACCGCGCCNGTTTCAGGCAAAATCGCCGCCATC...... 237
                      ||||||:::||||||   ::::::||||||
     81 ........ValHisAlaProThrSerGlyThrValThrAlaIleAlaPro 94

    238 ............CATCGCGGCGAAAAGCGCGTACTTCAGTCGGTCGTGAT 275
                    |||   :::            |||   ||||||:::||
     95 HisSerThrAlaHisProSerAlaLeuAlaGluLeu...SerValIleIl 110

    276 TGCCGTTGAAGGCAACGAC........................... 294
        |   ::::::|||:::|||
    110 eAspAlaAspGlyGluAspCysTrpIleProArgAspGlyTrpAlaAspT 127

    295 .................GAAATCGAGTTCGAACGCTACGCGCCCGAAGCG 327
                        |||      :::||||||            :::
    127 yrArgThrArgSerArgGluGluLeuIleGluArgIleHisGlnPheGly 143

    328 TTGGCAAACTTAAGCGGCGANGAANTNNGNNGCAATCTGATCCAATCCGG 377
        :::|||:::|||:::|||                          :::||
    144 ValAlaGlyLeuGlyGly.........................AlaGl 151

    378 TTTGTGGACTGCGCTGCGTANCCGTCCGTTCAGCAAAATCCCTGCCGTCG 427
        |:::   |||::::::                  |||:::   :::
    151 yPheProThrGlyVal.................LysLeuGlnGlyGlyG 162

    428 ATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATGGACACCAATCCGCTN 477
            :::      ::::::::::::|||||   :::   :::|||
    162 lyAspLysIleGluThrLeuIleIleAsnAlaAlaGluCysGluProTyr 178

    478 GCGGCAGACCCTGTGGTTGTGATCAAAGAAGCCGNCGANGATTTCAGACG 527
            :::          ::::::::::::::::::::::::      ::
    179 IleThrAlaAspAspArgLeuMetGlnAspCysAlaAlaGlnValValGl 195

    528 ANGTNTGCTGGTATTGAGCCGTTTGACCGAG................... 558
        :        :::|||:::::::::   :::
    195 uGlyIleArgIleLeuAlaHisIleLeuGlnProArgGluIleLeuIleG 212

    559 ....................CGTAAAATCCATGTGTGTAAGGCAGCTGGC 588
                            :::   |||   :::   :::|||:::
    212 lyIleGluAspAsnLysProGlnAlaIleSerMetLeuArgAlaValLeu 228

    589 GCAGACGTGCCGTCTGAAAATGCTGCCAACATCGAAACACATGAATTCGG 638
        ||||||                :::   :::|||:::   :::   :::
    229 AlaAsp..............SerAsnAspIleSerLeuArgValIlePr 240

    639 CGGCCCGCATCCGGCCGGT.....................TTGAGTG 664
              :::|||:::|||                       |||:::|
    240 oThrLysTyrProSerGlyGlyAlaLysGlnLeuThrTyrIleLeuThrG 257
```

```
      665 GCACGCACATTCATTTCATTGAGCCGGTCGGTGCAAACAAAACCGTTTGG 714
          ||   ::::::            |||   |||::::::::::::
      257 lyLysGlnVal............ProHisGlyGlyArgSerSerAspIle 269

      715 ACCATCAATTATCAAGATGTA.........ATTGCCATCGGACGTTTGTT 755
            :::      |||:::|||           |||:::   |||
      270 GlyValLeuMetGlnAsnValGlyThrAlaTyrAlaValLysArgAlaVa 286

      756 TGCAACAGGCCGTCTGAACACCGAGCGCGTGATTGCTTTGGGTGGTTCTC 805
                |||:::      ||||||||||||:::::|||   |||:::
      286 lIleAspGlyGluProIleThrGluArgValValThrLeuThrGlyGluA 303

      806 AAGTCAACAAACCACGCCTCTTGCGTACCGTTTTGGGTGCGAAAGTATCG 855
            :::   :::|||      :::   :::   ||||||:::   |||
      303 laIleAlaArgProGlyAsnValTrpAlaArgLeuGlyThrProValArg 319

      856 CAAATT......ACTGCGGGCGAATTGGTTGACGCAGACAACCGCGTGAT 899
          ::::::         ||||||         :::|||||||:::   |||||
      320 HisLeuLeuAsnAspAlaGlyPheCysProSerAlaAspGlnMetValIl 336

      900 TTCCGGTTCGGTATTGAACGGC 921
          |   |||:::   |||   |||
      336 eMetGlyGlyProLeuMetGly 343
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG00972

seq_documentation_block:

ID ABG00972 standard; Protein; 4274 AA.
XX
AC ABG00972;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #963.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS65159.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 31331; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4274 AA;

alignment_scores:

```
            Quality:  101.50              Length:    455
              Ratio:  0.505                 Gaps:     18
 Percent Similarity:  44.176   Percent Identity: 21.538
```

alignment_block:

US-09-303-518D-127/rev x ABG00972 ..

Align seg 1/1 to: ABG00972 from: 1 to: 4274

```
     1343 CAGCCTTCCTTCTCAANGGTTTCCAGCACCTTACGCAACAGCGGGCCNTA 1294
          :::|||   ||||||:::                    :::::::::|||
     1495 LysProPhePheSerThrArgProTyrGlnSerTrpThrThrAlaProIl 1511

     1293 TTCGTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGT 1244
           :::   :::||||||   :::|||          :::      |||||||
     1511 eThr...ValProGlyProAlaLysSerGlyPheThrSerLeuSerSerS 1527

     1243 CCAATTCCAAGCAACCCAATGCTTGCGCGCTG.............TCGGTA 1206
          ||:::|||:::   |||:::|||      |||            ||||||
     1527 erSerSerAsnThrProSerAlaSerProLeuLysSerIleTrpSerVal 1543

     1205 TCGCCGACGATTAAATCGCGCAAAAGCAGGGTAGGCAGGATGTCTAGCGG 1156
          |||   |||   :::      ||||||   :::|||      ::::::::
     1544 Ser...ThrProSerProIleLysSerThrLeuGlyAlaSerThrThrSe 1559

     1155 CATTACGCGCTCG..................................... 1143
          :   ::::::|||
     1559 rSerValLysSerIleSerAspValAlaSerProIleArgSerLeuArgT 1576

     1143 ..................................................... 1143

     1576 hrMetSerSerProIleLysThrValValSerGlnSerProTyrAsnIle 1592

     1142 TAAGTACCAATCGGCACCATGGCGCGGTCGCCACCGTTGACGGCTGTCGT 1093
             |||      ||||||:::||||||:::|||   :::|||   :::::
     1593 GlnValSerSerGlyThrLeuAlaArgAlaProAlaValThrGluAlaTh 1609

     1092 GAACTTGAAGAGTTTGTTTTTCAGGAAATGGCCGAGGGTCGTACGCGTGA 1043
          :   ||||||:::|||      ::::::
     1609 rProLeuLysGlyLeuAlaSerAsnSer..................... 1618

     1042 TGGAGTATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTG 993
               :::   |||      :::   :::         :::|||||||||
     1619 ..ThrPheSerSerArgThrSerProValThrThrAlaGlySerLeuLeu 1634

      992 CGGCCTTCTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATC 943
          :::   ||||||||||||
     1635 GluArgSerSerIleThr.............................. 1640

      942 GTGCGCGCCTTGTGTAATCGCGCCGTTCAATACCGAACCGGAAATCACGC 893
                     :::      |||   :::  :::   :::|||:::
     1641 .............MetThrProProAlaSerProLysSerAsnIleAsnM 1653

      892 GGTTGTCTGCGTCAACCAATTCGCCCGCAGTAATTTGCGATACTTTCGCA 843
               |||:::|||           ::::::|||   ::::::   |||
     1653 etTyrSerSerSerLeuProPheLysSerIleIleThrSerAla...Ala 1668
```

```
 842 CCCAAAACGGTACGCAAGAGGCGTGGTTTGTTGACTTGAGAACCACCCAA 793
     |||                          |||      :::|||   ||
1669 Pro.........................LeuIleSerSerProLeuLy 1676

 792 AGCAATCACGCGCTCGGTGTTCAGACGGCCTGTTGCAAACAAACGTCCGA 743
     |:::::::::      |||                 ::::::|||
1676 sSerValValSerProVal.................LysSerArgValA 1687

 742 TGGCAATTACATCTTGATAATTGATGGTCCAAACGGTTTTGTTTGCACCG 693
       :::|||:::|||      :::   :::   :::   |||   :::|||
1687 spValIleSerSerAlaLysIleThrMetAlaSerSerLeuSerSerPro 1703

 692 ....................................ACCGGCTCAAT 682
                                         :::|||||||::
1704 ValLysGlnMetProGlyHisAlaGluValAlaLeuValAsnGlySerIl 1720

 681 GAAATGAATGTGCGTGCCACTCAAACCGGCC................... 651
     ::::            |||||||||   |||
1720 eSer............ProLeuLysTyrAlaSerSerSerThrLeuIleA 1733

 650 ..GGATGC..............................GGGCCGCCG 636
       ||||||                            :::
1733 snGlyCysLysAlaThrAlaThrLeuGlnGluLysIleSerSerAlaThr 1749

 635 AATTCATGTGTTTCGATGTTGGCAGCATTTTCAGACGGC........... 597
     ||||||      |||:::::::::|||   :::|||
1750 AsnSerValSerSerValValSerAlaAlaThrAspThrValGluLysVa 1766

 596 .............................................A 596
                                                 :
1766 lPheSerThrThrThrAlaMetProPheSerProLeuArgSerTyrValS 1783

 595 CGTCTGCGCCAGCTGCCTTACACACATGGATTTTACGCTCGGTCAAACGG 546
     :::::||||||:::|||:::::::::           |||::::::
1783 erAlaAlaProSerAlaPheGlnSerLeuArgThrProSerAlaSerAla 1799

 545 CTCAATACCAGCANACNTCGTCTGAAATCNTCGNCGGCTTCTTTGATCAC 496
     |||   ||||||                        :::|||:::   ::
1800 LeuTyrThrSerLeu.................GlySerSerIleSerAl 1810

 495 AACCACAGGGTCTGCCGCNAGCGGATTGGTGTCCATC............. 459
     :||||||:::|||:::::::|||:::::::::::::::
1810 aThrThrSerSerValThrSerSerIleIleThrValProValTyrSerV 1827

 458 ......................GCATTGACGAAGATG......GCGAAC 438
                           ||||||   |||:::      :::|||
1827 alValAsnValLeuProGluProAlaLeuLysLysLeuProAspSerAsn 1843

 437 GGCTCGGCATCGACGGCAGGGATTTTGCTGAACGGACGGNTACGCAGCGC 388
     :::   :::::::::|||:::   ||||||:::                ::
1844 SerPheThrLysSerAlaAlaAlaLeuLeuSerProIleLysThrLeuTh 1860

 387 AGTCCACAAACCGGATTGGATCAGATTGCNNCNNANTTCNTCGCCGCTTA 338
     :::::::                 :::::::::   :::|||||||||:::|
1860 rThrGluThrHisProGlnProHisPheSerArgThrSerSerProValL 1877

 337 AGTTTGCC......AACGCTTCGGGCGCGTAGCGTTCGAACTCGATTTCG 294
     ||   :::         |||   :::|||   :::   ::::::   |||
1877 ysSerSerLeuPheLeuAlaProSerAlaLeuLysLeuSerThrProSer 1893

 293 TCGTTGCCTTCAACG 279
     ||||||   |||:::
1894 SerLeuSerSerSer 1898
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG07375

seq_documentation_block:

```
ID   ABG07375 standard; Protein; 4386 AA.
XX
AC   ABG07375;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #7366.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS71562.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 20; SEQ ID No 37734; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   4386 AA;
```

alignment_scores:

```
           Quality:  101.50            Length:    455
             Ratio:  0.505               Gaps:     18
 Percent Similarity:  44.176   Percent Identity: 21.538
```

alignment_block:

US-09-303-518D-127/rev x ABG07375 ..

Align seg 1/1 to: ABG07375 from: 1 to: 4386

```
1343 CAGCCTTCCTTCTCAANGGTTTCCAGCACCTTACGCAACAGCGGGCCNTA 1294
     :::|||   ||||||:::                    :::::::::|||
1504 LysProPhePheSerThrArgProTyrGlnSerTrpThrThrAlaProIl 1520

1293 TTCGTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGT 1244
       :::   :::|||||||   :::|||          :::      |||||||
1520 eThr...ValProGlyProAlaLysSerGlyPheThrSerLeuSerSerS 1536
```

```
1243 CCAATTCCAAGCAACCCAATGCTTGCGCGCTG.....TCGGTA 1206
     ||:::|||:::   |||:::|||      |||           ||||||
1536 erSerSerAsnThrProSerAlaSerProLeuLysSerIleTrpServal 1552

1205 TCGCCGACGATTAAATCGCGCAAAAGCAGGGTAGGCAGGATGTCTAGCGG 1156
     |||   |||   :::      ||||||   :::|||      ::::::::
1553 Ser...ThrProSerProIleLysSerThrLeuGlyAlaSerThrThrSe 1568

1155 CATTACGCGCTCG..................................... 1143
     :   ::::::|||
1568 rSerValLysSerIleSerAspValAlaSerProIleArgSerLeuArgT 1585

1143 ..................................................... 1143

1585 hrMetSerSerProIleLysThrValValSerGlnSerProTyrAsnIle 1601

1142 TAAGTACCAATCGGCACCATGGCGCGGTCGCCACCGTTGACGGCTGTCGT 1093
        |||      ||||||:::||||||:::|||   :::|||   ::::: 
1602 GlnValSerSerGlyThrLeuAlaArgAlaProAlaValThrGluAlaTh 1618

1092 GAACTTGAAGAGTTTGTTTTTCAGGAAATGGCCGAGGGTCGTACGCGTGA 1043
     :   ||||||:::|||      ::::::
1618 rProLeuLysGlyLeuAlaSerAsnSer..................... 1627

1042 TGGAGTATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTG 993
          :::   |||      :::   :::         :::|||||||||
1628 ..ThrPheSerSerArgThrSerProValThrThrAlaGlySerLeuLeu 1643

 992 CGGCCTTCTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATC 943
     :::   ||||||||||||
1644 GluArgSerSerIleThr.............................. 1649

 942 GTGCGCGCCTTGTGTAATCGCGCCGTTCAATACCGAACCGGAAATCACGC 893
               :::      |||   :::   :::   :::|||:::
1650 .............MetThrProProAlaSerProLysSerAsnIleAsnM 1662

 892 GGTTGTCTGCGTCAACCAATTCGCCCGCAGTAATTTGCGATACTTTCGCA 843
          |||:::|||           :::::::|||   ::::::   |||
1662 etTyrSerSerSerLeuProPheLysSerIleIleThrSerAla...Ala 1677

 842 CCCAAAACGGTACGCAAGAGGCGTGGTTTGTTGACTTGAGAACCACCCAA 793
     |||                        |||      :::|||   ||
1678 Pro.........................LeuIleSerSerProLeuLy 1685

 792 AGCAATCACGCGCTCGGTGTTCAGACGGCCTGTTGCAAACAAACGTCCGA 743
     |:::::::::      |||                ::::::|||
1685 sSerValValSerProVal.................LysSerArgValA 1696

 742 TGGCAATTACATCTTGATAATTGATGGTCCAAACGGTTTTGTTTGCACCG 693
       :::|||:::|||      :::   :::   :::   |||   :::|||
1696 spValIleSerSerAlaLysIleThrMetAlaSerSerLeuSerSerPro 1712

 692 .....................................ACCGGCTCAAT 682
                                          :::|||||||::
1713 ValLysGlnMetProGlyHisAlaGluValAlaLeuValAsnGlySerIl 1729

 681 GAAATGAATGTGCGTGCCACTCAAACCGGCC.................. 651
     ::::           |||||||||   |||
1729 eSer.....ProLeuLysTyrAlaSerSerSerThrLeuIleA 1742

 650 ..GGATGC................................GGGCCGCCG 636
       ||||||                                :::
1742 snGlyCysLysAlaThrAlaThrLeuGlnGluLysIleSerSerAlaThr 1758

 635 AATTCATGTGTTTCGATGTTGGCAGCATTTTCAGACGGC..........597
     ||||||      |||:::::::::|||   :::|||
1759 AsnSerValSerSerValValSerAlaAlaThrAspThrValGluLysVa 1775

 596 .....................................................A 596
                                                          :
1775 lPheSerThrThrThrAlaMetProPheSerProLeuArgSerTyrValS 1792

 595 CGTCTGCGCCAGCTGCCTTACACACATGGATTTTACGCTCGGTCAAACGG 546
```

```
          :::::||||||:::|||:::::::::           |||::::::
1792 erAlaAlaProSerAlaPheGlnSerLeuArgThrProSerAlaSerAla 1808

 545 CTCAATACCAGCANACNTCGTCTGAAATCNTCGNCGGCTTCTTTGATCAC 496
     |||   ||||||                          :::|||:::   ::
1809 LeuTyrThrSerLeu.................GlySerSerIleSerAl 1819

 495 AACCACAGGGTCTGCCGCNAGCGGATTGGTGTCCATC.............. 459
     :||||||:::|||:::::|||:::::::::::::::
1819 aThrThrSerSerValThrSerSerIleIleThrValProValTyrSerV 1836

 458 .......................GCATTGACGAAGATG......GCGAAC 438
                             ||||||   |||:::      :::|||
1836 alValAsnValLeuProGluProAlaLeuLysLysLeuProAspSerAsn 1852

 437 GGCTCGGCATCGACGGCAGGGATTTTGCTGAACGGACGGNTACGCAGCGC 388
     :::   :::::::::|||:::   ||||||:::                ::
1853 SerPheThrLysSerAlaAlaAlaLeuLeuSerProIleLysThrLeuTh 1869

 387 AGTCCACAAACCGGATTGGATCAGATTGCNNCNNANTTCNTCGCCGCTTA 338
     :::::::               :::::::::   :::||||||||||:::|
1869 rThrGluThrHisProGlnProHisPheSerArgThrSerSerProValL 1886

 337 AGTTTGCC......AACGCTTCGGGCGCGTAGCGTTCGAACTCGATTTCG 294
     ||   :::         |||   :::|||   :::   ::::::   |||
1886 ysSerSerLeuPheLeuAlaProSerAlaLeuLysLeuSerThrProSer 1902

 293 TCGTTGCCTTCAACG 279
     ||||||   |||:::
1903 SerLeuSerSerSer 1907
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG21944

seq_documentation_block:

ID ABG21944 standard; Protein; 4397 AA.
XX
AC ABG21944;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #21935.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS86131.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 52303; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

```
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   4397 AA;


alignment_scores:
           Quality:  101.50            Length:    455
             Ratio:  0.505               Gaps:     18
 Percent Similarity:  44.176   Percent Identity: 21.538

alignment_block:
 US-09-303-518D-127/rev x ABG21944   ..

 Align seg 1/1  to: ABG21944  from: 1  to: 4397

    1343 CAGCCTTCCTTCTCAANGGTTTCCAGCACCTTACGCAACAGCGGGCCNTA 1294
         :::|||   ||||||:::                   ::::::::::|||
    1504 LysProPhePheSerThrArgProTyrGlnSerTrpThrThrAlaProIl 1520

    1293 TTCGTATTTGCCCGGGCAGACGAAGCTGCACAAAGCGAGGTCTTCTTCGT 1244
          :::   :::||||||   :::|||         :::      |||||||
    1520 eThr...ValProGlyProAlaLysSerGlyPheThrSerLeuSerSerS 1536

    1243 CCAATTCCAAGCAACCCAATGCTTGCGCGCTG.....TCGGTA 1206
         ||:::|||:::   |||:::|||      |||           ||||||
    1536 erSerSerAsnThrProSerAlaSerProLeuLysSerIleTrpSerVal 1552

    1205 TCGCCGACGATTAAATCGCGCAAAAGCAGGGTAGGCAGGATGTCTAGCGG 1156
         |||   |||   :::      ||||||   :::|||      ::::::::
    1553 Ser...ThrProSerProIleLysSerThrLeuGlyAlaSerThrThrSe 1568

    1155 CATTACGCGCTCG..................................... 1143
         :   ::::::|||
    1568 rSerValLysSerIleSerAspValAlaSerProIleArgSerLeuArgT 1585

    1143 ..... 1143

    1585 hrMetSerSerProIleLysThrValValSerGlnSerProTyrAsnIle 1601

    1142 TAAGTACCAATCGGCACCATGGCGCGGTCGCCACCGTTGACGGCTGTCGT 1093
            |||      ||||||:::||||||:::|||   :::|||   :::::
    1602 GlnValSerSerGlyThrLeuAlaArgAlaProAlaValThrGluAlaTh 1618

    1092 GAACTTGAAGAGTTTGTTTTTCAGGAAATGGCCGAGGGTCGTACGCGTGA 1043
         :   |||||||:::|||      ::::::
    1618 rProLeuLysGlyLeuAlaSerAsnSer..... 1627

    1042 TGGAGTATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTG 993
              :::   |||      :::   :::         :::||||||||||
    1628 ..ThrPheSerSerArgThrSerProValThrThrAlaGlySerLeuLeu 1643

     992 CGGCCTTCTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATC 943
         :::   ||||||||||||
    1644 GluArgSerSerIleThr..... 1649

     942 GTGCGCGCCTTGTGTAATCGCGCCGTTCAATACCGAACCGGAAATCACGC 893

                  :::      |||   :::   :::   :::|||:::
    1650 .....MetThrProProAlaSerProLysSerAsnIleAsnM 1662

     892 GGTTGTCTGCGTCAACCAATTCGCCCGCAGTAATTTGCGATACTTTCGCA 843
              |||:::|||           ::::::|||   ::::::   |||
    1662 etTyrSerSerSerLeuProPheLysSerIleIleThrSerAla...Ala 1677

     842 CCCAAAACGGTACGCAAGAGGCGTGGTTTGTTGACTTGAGAACCACCCAA 793
         |||                              |||      ::::|||   ||
    1678 Pro.....LeuIleSerSerProLeuLy 1685

     792 AGCAATCACGCGCTCGGTGTTCAGACGGCCTGTTGCAAACAAACGTCCGA 743
         |:::::::::      |||                        :::::::|||
    1685 sSerValValSerProVal.....LysSerArgValA 1696

     742 TGGCAATTACATCTTGATAATTGATGGTCCAAACGGTTTTGTTTGCACCG 693
           :::|||:::|||      :::   :::   :::   |||   :::|||
    1696 spValIleSerSerAlaLysIleThrMetAlaSerSerLeuSerSerPro 1712

     692 .....ACCGGCTCAAT 682
                                                :::|||||||::
    1713 ValLysGlnMetProGlyHisAlaGluValAlaLeuValAsnGlySerIl 1729

     681 GAAATGAATGTGCGTGCCACTCAAACCGGCC..... 651
         ::::          |||||||||   |||
    1729 eSer.....ProLeuLysTyrAlaSerSerSerThrLeuIleA 1742

     650 ..GGATGC.....GGGCCGCCG 636
           ||||||          :::
    1742 snGlyCysLysAlaThrAlaThrLeuGlnGluLysIleSerSerAlaThr 1758

     635 AATTCATGTGTTTCGATGTTGGCAGCATTTTCAGACGGC..... 597
         ||||||      |||::::::::::|||   :::|||
    1759 AsnSerValSerSerValValSerAlaAlaThrAspThrValGluLysVa 1775

     596 .....A 596
                                                :
    1775 lPheSerThrThrThrAlaMetProPheSerProLeuArgSerTyrValS 1792

     595 CGTCTGCGCCAGCTGCCTTACACACATGGATTTTACGCTCGGTCAAACGG 546
         :::::|||||||:::|||:::::::::          |||::::::
    1792 erAlaAlaProSerAlaPheGlnSerLeuArgThrProSerAlaSerAla 1808

     545 CTCAATACCAGCANACNTCGTCTGAAATCNTCGNCGGCTTCTTTGATCAC 496
         |||   ||||||                          :::|||:::   ::
    1809 LeuTyrThrSerLeu.....GlySerSerIleSerAl 1819

     495 AACCACAGGGTCTGCCGCNAGCGGATTGGTGTCCATC..... 459
         :|||||:::|||:::::::|||:::::::::::::::
    1819 aThrThrSerSerValThrSerSerIleIleThrValProValTyrSerV 1836

     458 .....GCATTGACGAAGATG......GCGAAC 438
                                |||||   |||:::      :::|||
    1836 alValAsnValLeuProGluProAlaLeuLysLysLeuProAspSerAsn 1852

     437 GGCTCGGCATCGACGGCAGGGATTTTGCTGAACCGACGGNTACGCAGCGC 388
         :::   ::::::::::|||:::   ||||||:::                ::
    1853 SerPheThrLysSerAlaAlaAlaLeuLeuSerProIleLysThrLeuTh 1869

     387 AGTCCACAAACCGGATTGGATCAGATTGCNNCNNANTTCNTCGCCGCTTA 338
         :::::::               :::::::::   :::|||||||||:::|
    1869 rThrGluThrHisProGlnProHisPheSerArgThrSerSerProValL 1886

     337 AGTTTGCC......AACGCTTCGGGCGCGTAGCGTTCGAACTCGATTTCG 294
         ||   :::         |||   :::|||   :::   ::::::   |||
    1886 ysSerSerLeuPheLeuAlaProSerAlaLeuLysLeuSerThrProSer 1902

     293 TCGTTGCCTTCAACG 279
         ||||||   |||:::
    1903 SerLeuSerSerSer 1907

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR59843
```

```
seq_documentation_block:
ID   AAR59843 standard; Protein; 311 AA.
XX
AC   AAR59843;
XX
DT   26-JAN-1995  (first entry)
XX
DE   ApoE4Lx2 protease.
XX
KW   ApoE4Lx2; protease; enzyme; Alzheimer disease; diagnostic; therapeutic.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain
FT   Region          89..104
FT                   /label= apoE4 homology
FT   Cleavage-site   35..36
FT   Cleavage-site   43..44
FT   Cleavage-site   182..183
FT   Cleavage-site   295-296
FT   Misc-difference 3
FT                   /label= kinase c phosphorylation site
FT   Misc-difference 42
FT                   /label= kinase c phosphorylation site
FT   Misc-difference 43
FT                   /label= kinase c phosphorylation site
FT   Misc-difference 54
FT                   /label= kinase c phosphorylation site
FT   Misc-difference 87
FT                   /label= kinase c phosphorylation site
FT   Misc-difference 122
FT                   /label= kinase c phosphorylation site
FT   Misc-difference 129
FT                   /label= kinase c phosphorylation site
FT   Misc-difference 152
FT                   /label= kinase c phosphorylation site
FT   Misc-difference 303
FT                   /label= kinase c phosphorylation site
FT   Misc-difference 176..177
FT                   /label= casein kinase phosphorylation site
FT   Misc-difference 193..194
FT                   /label= casein kinase phosphorylation site
XX
PN   WO9413798-A.
XX
PD   23-JUN-1994.
XX
PF   16-DEC-1993;   93WO-EP03581.
XX
PR   16-DEC-1992;   92CA-2085924.
PR   04-MAR-1993;   93US-0291401.
XX
PA   (BERG/) BERGMANN J E.
PA   (PRED/) PREDDIE R E.
XX
PI   Bergmann JE,  Preddie RE;
XX
DR   WPI; 1994-234212/28.
DR   N-PSDB; AAQ69101.
XX
PT   New proteinase esterase-like proteins - used to develop prods.
PT   for the diagnosis and treatment of Alzheimer's disease and
PT   related diseases
XX
PS   Claim 26; Page 44-45; 72pp; English.
XX
CC   ApoE4Lx2 is a protease catalyzing the formation of the abnormal beta/A4
CC   variant of beta-amyloid protein, and is used to develop an inhibitor
CC   for the diagnosis and teatment of Alzheimer disease, Downs syndrome,
CC   Parkinson disease, schizophrenia, hyperlipoproteinemia or
CC   cardiovascular disease.
XX
```

```
SQ   Sequence   311 AA;


alignment_scores:
            Quality:  101.00              Length:    318
              Ratio:  0.754                 Gaps:     15
 Percent Similarity:  42.138   Percent Identity: 24.214

alignment_block:
 US-09-303-518D-127/rev x AAR59843   ..

 Align seg 1/1  to: AAR59843  from: 1  to: 311

    1268 CTGCACAAAGCGAGGTCTTCTTCGTCCAATTCCAAGCAACCCAATGCTTG 1219
         |||::::::|||   ::::::|||:::      :::        |||||
      11 LeuArgArgAlaTrpLysAlaSerAlaCysArgArgIleCysTrpAlaCy 27

    1218 CGCGCTGTCGGTATCGCCGACG.................ATTAAATCGC 1187
         |:::   |||::::::   |||                    :::||||
      27 sSerSerSerLeuAlaArgThrSerAlaThrCysSerPheThrSerSerA 44

    1186 GCAAAAGCAGGGTAGGCAGGATGTCTAGCGGCATTACGCGCTCGTAAGTA 1137
         ||:::|||||||||          ::::::::::|||||||||:::
      44 rgArgSerArgValArgLeuPro.IleSerSerIleArgAlaArgSerAr 60

    1136 CCAATCGGCACCATGGCGCGGTCGCCACCGTTGACGGCTGTCGTGAACTT 1087
          :::         |||
      60 gSerProGlnAlaTrpAla.............................. 66

    1086 GAAGAGTTTGTTTTTCAGGAAATGGCCGAGGGTCGTACGCGTGATGGAGT 1037
           |||   |||   ||||||                          |||
      67 ..ArgSerCys...SerGly.......................Trp... 72

    1036 ATTTGTCCGGCTGCGGCGCAACCCAGCCGAACAGCTCTTTGCTGCGGCCT 987
                    |||   :::|||:::   :::|||
      73 .........ProAlaArgGluProThrValAlaAla.............. 81

     986 TCTTCGATAACGGAAATCTGATTGTGGTAGCGTCCCAAATAATCGTGCGC 937
                 |||         |||::::::         :::|||:::|
      82 ..........ArgThrArgProCysSerThrArgGlyProArgArgSerA 95

     936 GCCTTGTGTAATCGCGCCGTTCAATACCGAACCGG.............. 902
         ||::::::         ||||||   |||:::|||
      95 rgMetAlaLeuArgProArgSerAlaProSerArgAlaProAlaTrpTyr 111

     901 ...AAATCACGCGGTTGTCTGCGTCAACCAATTCGCCCGCAGTAATTTGC 855
                 |||   |||      :::      ||| ||||||:::||||
     112 ThrAlaArgArgPheCysArgSerSerAlaSerArg.ArgSerArgLeuA 128

     854 GATACTTTCGCACCCAAAACGGTACGCAAGAGGCGTGGTTTGTTGACTTG 805
         ||   :::|||                        |||      :::
     128 rgSerLeuArg.....................ArgTrp......GluAla 135

     804 AGAACCACCCAAAGCAATCACGCGCTCGGTGTTCAGACGGCCTGTTGCAA 755
         |||   |||:::|||:::   :::|||      :::   ||||||   ::
     136 ArgArgThrArgSerSerSerValLeuTrpProSerMetAlaCysThrSe 152

     754 ACAAACGTCCGATGGCAATTACATCTTGATAATTGATGGTCCAAACGGTT 705
         :
     152 rProArg.......................................... 154

     704 TTGTTTGCACCGACCGGCTCAATGAAATGAATGTGCGTGCCACTCAAACC 655
             ||||||                        ||||||   :::|||
     155 ..TyrCysThr.....................ArgAlaProArgThr 162

     654 GGCCGGATGCGGGCCGCCGAATTCATGTGTTTCGATGTTGGCAGCATTTT 605
         :::   |||   |||   :::      |||         :::|||:::
     163 SerSerMetSerAlaProSerArgAlaCysAlaAlaCysSerSerLeuAs 179

     604 CAGACGGCACGTCTGCGCCAGCTGCCTTACACACATGGATTTTACGCTCG 555
             |||         |||||||:::
     179 pSerArgAlaArgValAlaSerAla..................... 187
```

```
554 GTCAAACGGCTCAATACCAGCANACNTCGTCTGAAATCNTCGNCGGCTTC 505
        |||:::          :::::::|||||||:::::::::::::
188 .....ThrGlyValSerCysSerSerSerSerAspLeuAlaAlaPheAsn 202

504 TTTGATCACAACCACAGGGTCTGC.............CGCNAGCGGATTGG 467
      :::                 :::||||          |||   |||   ||
203 SerSerMetValSerSerIleCysAlaValArgProArgGlyArgGluGl 219

466 TGTCCATCGCATTGACGAAGATGGCGAACGGCTCGGCATCGACGGCAGGG 417
    ||||   :::           ||||||   ||||||      |||||||
219 yValAlaGlnTrpGlyGlyGlyGlyGluGlyLeuGlyTrpGlyGlyArgA 236

416 ATTTTGCTGAACGGACGGNTACGCAGCGCAGTCCACAAACCGGATTGGAT 367
      ||     |||             :::|||:::   ::::::|||   :::
236 spGlyAspGluProGluArgProLysArgGluArgArgAsnGlyAlaGlu 252
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG80035

seq_documentation_block:

```
ID   AAG80035 standard; Protein; 507 AA.
XX
AC   AAG80035;
XX
DT   17-JAN-2002  (first entry)
XX
DE   Pseudomonas putida KIE171-BII ipuH protein.
XX
KW   L-alaninol; isopropylamine; iupI; ipuH; ofloxacin; ipuC; glutamylamide;
KW   gamma-glutamylamide synthase; theanine; ipuH.
XX
OS   Pseudomonas putida.
XX
PN   WO200173038-A2.
XX
PD   04-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-EP03651.
XX
PR   31-MAR-2000; 2000EP-0106888.
XX
PA   (LONZ ) LONZA AG.
XX
PI   Leisinger T,  Van Der Ploeg J,  Kiener AM,  De Azevedo Waesch SI;
PI   Maire T;
XX
DR   WPI; 2001-626266/72.
DR   N-PSDB; AAI68687.
XX
PT   New microorganism for converting isopropylamine to L-alaninol, useful
PT   as pharmaceutical intermediate especially ofloxacin, has the ipuH and I
PT   genes for alaninol catabolism inactivated  -
XX
PS   Example 1; Page 105-106; 106pp; German.
XX
CC   This invention describes novel microorganisms (A), or their enzymatic
CC   extracts, able to convert isopropylamine (I) to L-alaninol (II) and which
CC   have the ipuH and ipuI genes, encoding enzymes involved in metabolism of
CC   (II), inactivated. (II) is an intermediate for pharmaceuticals,
CC   especially ofloxacin. Also the enzyme encoded by the bacterial ipuC gene
CC   (a gamma-glutamylamide synthase) is used for synthesis of gamma-
CC   glutamylamides, specifically theanine. Inactivation of the ipuH and I
CC   genes improves production of (II). Typically, a Pseudomonas in which both
CC   ipuH and I genes were inactivated produced (II) at 8 mM after 60 hour
CC   culture. For a similar strain in which only the ipuI had been inactivated
CC   the maximum (II) concentration was about 7 mM, after 6 hour, falling to
CC   practically zero after about 20 hour. This sequence encodes the
CC   Pseudomonas putida KIE171-BII IpuH protein product (potential aldehyde
CC   dehydrogenase) which is described in the method of the invention.
XX
SQ   Sequence   507 AA;
```

alignment_scores:

```
                Quality:  101.00          Length:    463
                  Ratio:  0.474             Gaps:     28
Percent Similarity:  46.004   Percent Identity: 22.246
```

alignment_block:
US-09-303-518D-127 x AAG80035 ..

Align seg 1/1 to: AAG80035 from: 1 to: 507

```
 64 GACGGGCCCGTCATTACCGAAGTCGCGTTGCTTGGCGAAGAATATGCCGG 113
    ||||||   |||:::|||:::|||:::      ||||||
 54 AspGlyArgValLeuThrLysValSerAlaGlyGlyGlu..... 66

114 TATGCGCCCCTNGATGAAAGTCAAGGAAGGCGATGCCGTCAAAAAAGGCC 163
                    :::     :::   ||||||      ::::
 67 ..................AlaAspValAsnArgAlaValAlaAlaAlaA 77

164 AAGTGCTGTTTGAAGACAAAAAGNATCCGGGCGTGGTGTTTACCGCGCCN 213
    ::      |||:::|||:::::      |||              ::|||
 77 rgSerAlaPheArgAspArgArgTrpSerGly.........GlnSerPro 90

214 GTTTCAGGCAAA.....................ATCGCCGCCATCCA 239
    ||||||   |||                     :::      :::||
 91 ValSerArgLysArgThrLeuGlnAlaPheAlaAlaLeuIleArgLeuHi 107

240 TCGCGGCGAAAAGCGCGTACTTCAGTCGGTC.....GTGATTG 277
    ||||   |||      :::|||:::::::::          |||:
107 sArgAspGluLeuAlaLeuLeuGluThrLeuAspMetGlyLysProIleS 124

278 CCGTTGAAGGCAACGACGAAATCGAG.............TTCGAACGC 312
    ::::::::   :::   ::::::|||              ||||||
124 erAlaSerArgSerValAspValGluAlaValAlaSerCysPheGluTrp 140

313 TACGCGCCCGAAGCGTTGGCAAACTTAAGCGGCGANGAANTNNGNNGCAA 362
    ||||||   |||||||:::   :::|||       |||        :::
141 TyrAla...GluAlaIleAspLysLeu......TyrGluGlnIleAlaPr 154

363 TCTGATCCAATCCGGTTTGTGGACTGCGCTGCGTANCCGTCCGTTCAGCA 412
          ::::::   |||   :::   :::      :::|||::::::
154 oThrAlaGluThrAspLeu...AlaLeuIleThrArgGluProLeuGlyV 170

413 AAATCCCTGCCGTCGATGCCGAGCCGTTCGCCATCTTCGTCAATGCGATG 462
      :::   |||:::                |||   ::::::::   |||
170 alValGlyAlaIleValProTrpAsnPheProMetLeuThrAlaAlaTrp 186

463 GACACCAATCCG...CTNGCGGCAGACCCTGTGGTTGTGATCAAAGAAGC 509
      |||   |||   |||||:::          ||||||:::|||   ||
187 LysThrAlaProAlaLeuAlaCysGlyAsnSerValValLeuLysProAl 203

510 CGNCGANGATTTCAGACGANGTNTGCTGGTATTGAGCCGTTTGACCGAGC 559
    |           :::::::::   |||:::      ||||||::::::
203 a............GluGlnSerProLeuThrAlaIleArgLeuAlaGln. 215

560 GTAAAATCCATGTGTGTAAGGCAGCTGGCGCAGACGTGCCGTCTGAAAAT 609
                        |||      |||   ||||||   |||:::
216 .................LeuAlaLeuGluAlaGlyValPro...GluGly 225

610 GCTGCCAACATCGAAACACATGAATTCGGCGGCCCGCATCCGGCCGGTTT 659
    :::   |||:::                              |||    ::|||
226 ValPheAsnValVal.....................ProGlyLeuGlyGlySe 236

660 GAGTGGCACGCACATTCATTTCATTGAGCCGGTCGGTGCAAACAAAACCG 709
    :::|||                    :::   ::::::::::::      |
236 rAlaGly.....................ArgAlaLeuAlaCysHisMetAspV 247

710 TTTGGACCATCAATTATCAAGATGTAATTGCCATCGGACGTTTGTTTGCA 759
    ||      |||   :::                |||   ||||||||::::::
247 alAspGlyIlePhePheThrGlySerThrAlaThrGlyArgLeuLeuThr 263

760 ......ACAGGCCGTCTGAACACCGAGCGCGTG...ATTGCTTTGGGTGG 800
          :::::::::   |||   ::::::::   :::   ||||||||
264 GluTyrAlaAlaLysSerAsnLeuLysArgValCysLeuGluLeuGlyGl 280
```

```
 801 TTCTCAAGTCAACAAACCACGCCTCTTGCGTACCGTTTTGGGT..... 843
 280 y.........LysSerProAsnIleIleLeuAlaSerTyrGlyAspIleG 294

 844 .....GCGAAAGTATCGCAAATTACTGCG.............GGCGAA 873
 294 luLysAlaAlaValAlaSerAlaGluSerMetPheAsnAsnGlnGlyGlu 310

 874 TTGGTTGACGCAGACAACCGCGTGATTTCCGGTTCGGTATTGAACGGCGC 923
 311 ValCysIleAlaProSerArgLeuIle..................... 319

 924 GATTACACAAGGCGCGCACGATTATTTGGGACGCTACCACAATCAGATTT 973
 320 .ValGluArgSerIleHisLysArgVal.................ValG 330

 974 CCGTTATCGAAGAAGGCCGCAGCAAAGAGCTGTTCGGCTGGGTTGCGCCG 1023
 330 luIleValAlaGluValAlaLysArgArg..................... 339

1024 CAGCCGGACAAATACTCCATCACGCGTACGACCCTCGGCCATTTCCTG.. 1071
 340 GlnProGlyAspProLeuAspProThrThrArgMetGlyAlaLeuValAs 356

1072 .........AAAAACAAACTCTTCAAGTTC..................... 1092
 356 pAlaAsnHisAlaAspArgValMetGlyPheIleGlyArgAlaLysAlaA 373

1093 .....ACGACAGCCGTCAACGGTGGCGACCGCGCCATGGTGCCGATT... 1134
 373 spGlyAlaThrLeuValAlaGlyGlyThrArgAlaLeuThrGluThrGly 389

1135 GGTACTTACGAGCGCGTAATGCCG.......................... 1158
 390 GlySerTyr......ValValProThrValPheAspAsnValSerAsnCy 404

1159 ..................................CTAGACATCCTGC 1171
 404 sMetGluIleAlaArgAspGluValPheGlyProValLeuSerValIleP 421

1172 CTACCCTGCTTTTGCGCGATTTAATCGTCGGCGATACCGACAGCGCGCAA 1221
 421 roValAlaAsnValGlyGluAlaValAlaValAlaAsnAspSerProTyr 437

1222 GCATTGGGTTGCTTGGAATTGGACGAAGAAGACCTCG 1258
 438 GlyLeuGly.AlaGlyValTrpThrAspArgLeuSer 449
```